ESD

PHYSICS AND DEVICES

STEVEN H. VOLDMAN

WILEY

ESD

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

ESD
Physics and Devices

Steven H. Voldman
Vermont, USA

John Wiley & Sons, Ltd

 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging in Publication Data

ESD: physics and devices / Steven H. Voldman.
p. cm.
Includes bibliographical references and index.
ISBN 0-470-84753-0 (cloth : alk. paper)
1. Electronic apparatus and appliances—Protection. 2. Electric discharges.
3. Electrostatics. I. Title.

TK7870.V567 2004
621.3815—dc22

2004010317

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-84753-0

Typeset in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

To My Wife Annie

and

To My Children

Rachel Pesha and Aaron Samuel

Contents

About the Author

Dr Steven H. Voldman received his B.S. in Engineering Science from the University of Buffalo (1979); M.S. EE (1981) and Electrical Engineer Degree (1982) from M.I.T; MS Engineering Physics (1986) and Ph.D EE (1991) from the University of Vermont under IBM's Resident Study Fellow program. At M.I.T, he worked as a member of the M.I.T. Plasma Fusion Center, and the High Voltage Research Laboratory (HVRL). At IBM, as a reliability/device engineer, his work included pioneering work in bipolar/CMOS SRAM alpha particle and cosmic ray SER simulation, MOSFET gate-induced drain leakage (GIDL) mechanism, hot electron, epitaxy/well design, CMOS latchup, and ESD. Since 1986, he has been responsible for defining the IBM ESD/latchup strategy for CMOS, SOI, BiCMOS and RF CMOS and SiGe technologies. He has authored ESD and latchup publications in the area of MOSFET scaling, device simulation, copper, low-k, MR heads, CMOS, SOI, SiGe and SiGeC technology. Voldman served as SEMATECH ESD Working Group Chairman (1996–2000), ESD Association General Chairman and Board of Directors, International Reliability Physics (IRPS) ESD/Latchup Chairman, International Physical and Failure Analysis (IPFA) Symposium ESD Sub-Committee Chairman, ESD Association Standard Development Chairman on Transmission Line Pulse Testing, ESD Education Committee, and serves on the ISQED Committee, Taiwan ESD Conference (T-ESDC) Technical Program Committee. Voldman has provided ESD lectures for universities (e.g. MIT Lecture Series, Taiwan National Chiao-Tung University, and Singapore Nanyang Technical University). He is a recipient of over 125 US patents, over 100 publications, and also provides talks on patenting, and invention. He has been featured in *EE Times, Intellectual Property Law and Business* and authored the first article on ESD phenomena for the October 2002 edition of *Scientific American* entitled *Lightning Rods for Nanostructures*, and *Pour La Science, Le Scienze, and Swiat Nauk* international editions. Dr. Voldman was recently accepted as the first IEEE Fellow for ESD phenomena in semiconductors for 'contributions to electrostatic discharge protection in CMOS, SOI and SiGe technologies.'

Preface

Electrostatic discharge (ESD) phenomena have been known to mankind since Thales of Miletus in approximately 600 B.C.E. noticed the attraction of strands of hay to amber. Two thousand six hundred years have passed and the quest to obtain a better understanding of electrostatics and ESD phenomenon continues. Today, the manufacturing of microelectronics has continued the interest in the field of electrostatic phenomenon spanning factory issues, tooling, materials, and the microelectronic industry.

Not only is the quest for the scientific understanding of ESD, there is a struggle in the way that the subject is taught and presented. At this time, there is only one university course in the world that teaches a semester course in the discipline of ESD phenomena in semiconductor components. Today, there is a dilemma of how to teach and educate engineers for ESD protection in components at university level to establish a ESD discipline. To date, ESD is taught in individual lectures, short courses, and tutorials. There are books on the subject, but they are not structured for formal undergraduate or graduate courses suited for physicists, mathematicians, materials science majors, or electrical engineers from a generalist perspective which would engage wide interest across the materials and semiconductor community.

The nature of ESD phenomena makes it a difficult subject to teach unless it is taught from a cross-disciplinary perspective or as a series of disciplines which are woven together carefully to build understanding from first principles. This same dilemma occurred in the early 1960s in the teaching of integrated semiconductor electronics. The Semiconductor Electronic Education Committee (SEEC) was formed to address how to teach an engineer integrated electronics in a world that separated the teaching of solid state physics, devices, and circuits. It was stated in the Foreword of the SEEC series: 'the development of micro-miniaturization of electronic circuits has blurred the dividing line between the "device" and the "circuit" and thus has made it increasingly important for us to understand deeply the relationship between internal physics and structure of a device, and its potentialities for circuit performance.' It was at this juncture, that the initiative to establish a series of semiconductor books which integrated understanding from the fundamentals of semiconductor physics to circuits in a coherent fashion, where each book built on the previous book with a bottom-up approach. In the ESD discipline, this same dilemma holds true. Yet it is an even more

difficult problem because one has the device, the circuit, the package, the system – and electrical, thermal and mechanical issues all at the same time. Additionally, the time scale response changes with the scale of interaction and the electrostatic phenomenon itself.

In the semiconductor industry, the perception of ESD phenomena was regarded as a 'black box' which appeared to have no basis of understanding or educational discipline behind the technical achievements. ESD engineers were viewed as semiconductor alchemists, providing solutions without the fundamental understanding or the ability to explain the achievements. The majority of the accomplishments were achieved by brute force or nonsystematic experimental procedures to achieve the results. ESD engineers ran on their educational background, on-the-job ESD learning and established intuition without the assistance of the proper tools to achieve their results.

In my own personal experience, engineers and management had no confidence in the ESD discipline – yet still needed to run the business, get out the designs, and ship products. This made me realize that the perception was that there had been no science or structure put in place in the past, and the progress made in the ESD discipline was not standing on firm ground. The common belief was there was no discipline in the ESD discipline. I believe part of the perception was due to the lack of a structured educational framework. Another observation was there was no common design practice, no common test structure, no standardization, no technology benchmarking, or ESD technology roadmap. This lack of strategic planning was partly based on the lack of strategic thinking, but also on the lack of predictive capability, or even the *belief* in predicting ESD trends and directions as semiconductor scaling occurred. Engineers did not even believe in the relationship between technology scaling and ESD predictive trends. These observations made me realize the lack of confidence in ESD discipline as a technical discipline or science – again pointing to the lack of formal understanding and cross-discipline development. Additionally, these concepts were reinforced since there was no undergraduate or graduate university course.

This motivated me to move forward with the concept of not a single book, but a series of books on electrostatic discharge phenomena. The objective of the book series is to establish an educational framework to establish an ESD discipline based on integration of physics, devices and circuits from the bottom upward. This will be a difficult task.

In this textbook, a first goal is to focus on ESD in semiconductor devices and circuits, yet to provide a text which is valuable and comprehensible to non-ESD engineers who have had course work in mechanics, thermal physics, mathematics, circuit design or semiconductor device physics. The goal of the text would be to provide educational value to ESD engineers who are not semiconductor device engineers. To achieve this goal, the material in the text must contain physics fundamentals which are both simplistic and classical. A focus on mathematics, and models, is key to establish the base of the ESD discipline. The hope is this will serve as a learning vehicle and reference source for the future.

A second goal is to connect early fundamentals to today's technology and future technologies. An objective of this text is to provide material which not only addresses today's problems, but will have long-term retention in its concepts. For this reason, literature containing physics-based models will be used to discuss today's technology concerns. In such a manner, the technology twists and turns along scaling roadmaps will be apparent. Additionally, new ESD disciplines will be shown to be rooted in the past

with connections to existing models and literature. By writing in this fashion, the fundamental concepts will be threaded and woven together.

A third goal is to provide depth as well as breadth in ESD engineering, especially in new technologies and new endeavors. Where significant materials exist, depth is avoided when other sources are available. On new subjects, some penetration into the materials is necessary to understand the subject as well as its significance.

A fourth goal is to provide an educational base which establishes qualitative understanding, intuition and inventiveness. At this time, prediction of ESD protection results is not a closed solution. This is not because that ESD analysis is intractable, but because ESD physics and product results both involve microscopic and macroscopic scales, and models or tools which are not extended over wide current, voltage and temperature ranges. The ESD 'models' and reality in actual product implementation are not connected by an exact one-to-one mapping. Being an ESD engineer requires both an intuition of why things are occurring, and being able to extrapolate that incomplete knowledge into a working implementation. Teaching intuition is not easy; teaching inventiveness is easier.

This first book in this series contains the following:

Chapter 1 introduces the reader to the time constants and physical pulse models associated with ESD. Understanding physical time constants allows one to determine which physical phenomenon is important. While the student is learning the fundamentals of the different known ESD 'models', the text provided a sense of the physical time constants of interaction. The time constants that define electroquasistatic (EQS), magnetoquasistatic (MQS), electromagnetic (EM), thermal diffusion processes, ESD pulse models, circuit time constants, and package models are key to understanding and intuition. Normalization, and dimensionless groups are also key to understanding the scale and level of the physical system. Chapter 1 focuses on concepts of stability and instability from an electrical and electrothermal perspective. Early work of Paschen, Toepler, and Townsend is discussed for the understanding of breakdown phenomenon. The chapter focuses on breakdown phenomenon in air, and in silicon to understand the avalanche phenomenon. Breakdown instabilities are discussed from both a time and spatial perspective which will be used in future discussion for understanding of MOSFETs and current constriction, and ballasting. From this approach, understanding of operation of everything from testers, spark gaps, ESD networks, and MOSFET snapback can be understood.

Chapter 2 focuses on thermal models, thermal physics and ESD electrothermal models. The majority of the fundamental electrothermal ESD models are rooted in the solutions to the heat diffusion equation under geometrical constraints, and time constant assumptions. The focus of this chapter will show general methods from the solutions in time and space, in spherical, cylindrical and rectangular geometries. The chapter will introduce mathematical techniques: Green's function solutions, method of images, Duhamel's formulation, Boltzmann and Kirchoff transformations, the flux-potential transfer matrix approach, stratified media, and other methods. This naturally leads to the early ESD models proposed by Wunsch and Bell, Tasca, Arkhipov, Dwyer and Campbell, Ash, Smith-Littau, and others. This teaching leads to demystifying the ESD models to simple realizations of thermal solutions and the realization of physical time constants. The magic is in the beauty of the mathematics of the thermal diffusion equation and how to apply it to simple geometrical structures and geometrical models based on a thermal time constant assumption!

Chapter 3 focuses on semiconductor devices and ESD issues. To avoid being a book on semiconductors, the chapter focuses on only certain areas of interest and devices of interest. The devices include diodes, resistors, bipolar transistors, silicon-controlled rectifiers, and MOSFETs. Without being exhaustive, we address high-current, high-voltage and high-temperature effects. The chapter addresses leakage phenomena from thermal models, trap-to-band models, band-to-band tunneling models which are important for understanding leakage. For resistor elements, the focus is on velocity saturation effects. For bipolar transistors, we discuss limitations such as the Johnson Limit, Kirk effect and basic models. For the silicon-controlled rectifier, the focus is on the criteria for triggering for different cases, and the general tetrode model. Holding current formulations will also be discussed. For MOSFETs, the focus is on MOSFET current constriction, avalanche phenomenon, parasitic bipolar, and MOSFET gate-induced drain leakage phenomenon. The chapter also addresses velocity saturation effects and how they relate to ESD phenomenon. The aim is to relate some of the concepts to the previous chapters, and also to future chapters.

In the subsequent chapters, each region of the semiconductor is focused on in a general manner, as opposed to creation of specific device models. The standard practice of teaching these effects is to apply it to a MOSFET transistor, or a diode, or a resistor. We generalize the subject, under the realization that the models can be integrated into all devices.

Chapter 4 focuses on substrates and ESD phenomena. Substrate electrical and thermal modeling influences all circuits and ESD elements. The general concepts and models can be applied to the diode, the bipolar transistor, the MOSFET, and the silicon-controlled rectifier. Teaching from the generalist approach, the substrate is a semi-infinite domain containing a plurality of devices, which can be treated independently or coupled. Substrate model concepts such as semi-infinite domain models, to transmission line models, and lossy transmission lines, are discussed. This is followed by electrical and thermal models using Green's function methods and new geometrical models. Additionally, to address multi-finger or a plurality of elements, the transfer resistance concept, used by Ron Troutman for latchup, is shown for more complicated scenarios and coupling phenomena. These models can be used for ESD, noise or latchup. Another approach is viewing the substrate as a stratified media. This technique and the flux-potential transfer matrix approach are used widely at MIT to address problems by J. Melcher in continuum electromechanics, and by J. A. Kong in electrodynamics for problem solving. These methods are applicable for problem solving in the ESD area for variable doped regions, SOI wafers to GaAs substrates.

Chapter 5 discusses well regions, and ESD. It addresses the distinctions between diffused wells, retrograde wells, triple well and sub-collectors from a general approach. Distinctions are based on ESD experimental work, showing that these have significant effect on the operation of vertical diodes, vertical parasitic *pnp* transistors, lateral *npn* transistors and other ESD structures of interest. This knowledge can be used for an understanding of ESD, latchup and minority-carrier injection. Samples of some experimental results will be shown of recent discoveries in ESD associated with well structures.

Chapter 6 discusses the influence of isolation structures on ESD robustness. Isolation structures include LOCOS isolation, shallow trench isolation and deep trench isolation. These structures have significant impact on the operation of ESD networks and latchup.

Distinctions are made between how ESD devices respond between LOCOS and STI. Simulation, and experimental results of LOCOS and STI structures are discussed.

Chapter 7 discusses MOSFET drain structures and salicides. MOSFET drain structures and salicide have significant influence on ESD protection. MOSFET drain engineering evolution as well as the transition from titanium to cobalt salicide will be discussed. A discussion on titanium properties, and transition to low resistance and the usage of molybdenum in assisting of the transition is briefly addressed. A focus is maintained on thermal models on salicides relating to the self-heating effects from ESD events.

Chapter 8 addresses dielectrics and ESD events. The model of Fong and Hu of dielectric failure is discussed. This is followed by the work of D. Lin, addressing the distinction of the model between d.c. and pulsed phenomenon. The chapter primarily discusses the physics of dielectrics. To address dielectrics in general would be quite exhaustive, hence a small view of the dielectrics is shown.

Chapter 9 addresses interconnects and ESD. Interconnects and ESD are discussed by addressing aluminum interconnects, copper interconnects, and copper/low-k inter-level dielectric systems. Although there is a number of interconnect models, the focus is on those proposed by Banerjee and the author. Models for resistors are suitable for interconnects, such as the Smith–Littau model. Additionally, models have been proposed by Maloney, Vinson, and Salome. The model discussed in the chapter on *n*-wells can be utilized as well for interconnects. Instead of an exhaustive list of interconnect models, new issues such as low-k dielectrics, fill shapes, voiding, ESD electromigration studies and other issues are briefly addressed. Materials from P. Y. Tan and S. Sherry on material changes and EM impacts are shown. These models are also suitable and applicable for understanding of ESD failures in magnetic recording devices such as magnetoresisitive heads, giant magnetoresistive heads, and tunneling magnetoresistive heads.

Chapter 10 addresses silicon on insulator (SOI) technology and ESD. SOI technology is arousing growing interest; it introduces new issues for ESD protection and is a natural extension of CMOS Technology. We discuss lateral polysilicon gated diode structures in SOI. SOI diode models, and thermal models are introduced. From the generalist approach, SOI modeling is a natural for Green's function thermal models due to the geometries, and the substrate can be treated as a stratified medium problem. The dynamic threshold SOI devices is also introduced to show new opportunities in SOI environments. The interest in SOI technology continues for both earth-based and space applications. This chapter shows the early work of SOI in modern CMOS technology. This chapter carries the reader from the first work at IBM starting in 1994 to the present day.

Chapter 11 covers silicon–germanium (SiGe) and silicon–germanium–carbon (SiGeC) technologies and ESD. Silicon–germanium device physics, and ESD phenomena are briefly discussed. The chapter focuses on the distinctions between silicon bipolar versus silicon–germanium, and the silicon–germanium versus silicon germanium carbon devices from an ESD perspective. It was as recent as 2000 that the first paper on ESD in SiGe bipolar transistors was published. This chapter brings the reader up to speed from the first publication to the present.

Chapter 12 introduces devices of the future where there are currently no reported ESD measurements! This chapter briefly addresses strained silicon devices, Fin-FETs, carbon nanotubes and other things to look forward to in the future.

Two millennia and six hundred years have passed since the discovery of electrostatic phenomena and it is only now that we have structured a text to evaluate the teaching of this subject in this fashion. In this text, I needed to find a happy medium which bridged between fundamentals and modern applicability to today's technology, and aimed to bring the reader from the past to what is going to happen tomorrow.

Enjoy the text, and enjoy the subject matter of ESD. There is still so much to learn.

B"H
Steven H. Voldman
IEEE Fellow

Acknowledgements

I would like to thank the individuals who have helped me in my academic and professional career path. There have many people in my past who have guided me, from a few words, to years of interaction. Faculty from the University of Buffalo and Massachusetts Institute of Technology had significant impact on my direction and interest in the area of continuum mechanics, continuum electromechanics, electrostatics and semiconductors. I am indebted to the Engineering Science curriculum at the University of Buffalo and to Professor Irving Shames, Professor Herbert Reismann and Professor Stephen Margolis for support and interest in the areas of thermal, mechanical, and electrical sciences. At Massachusetts Institute of Technology, I am indebted to the electrical engineering (EE) department, M.I.T. plasma fusion center, and M.I.T. high-voltage research laboratory (HVRL) for the support in the area of plasma physics, electrodynamics, electrostatics, and semiconductors. Professor James R. Melcher, Professor Markus Zahn, Professor Cliff Fonstad, Professor Louis D. Smullin, Professor J. A. Kong, Professor David Epstein and Professor C. Cook as well as other EE faculty contributed to my understanding of continuum electromechanics, electrostatics and semiconductors in the area of technical material, course material development, education and educational texts.

At IBM, I was fortunate to have many mentors, and friends from IBM Burlington Vermont, IBM East Fishkill, IBM T. J. Watson Research Center, as well as circuit designers from IBM Poughkeepsie, Rochester, Austin, Manassas, Kingston, Endicott and Raleigh. I have been fortunate for the last 22 years to have the advice and counsel of many IBM engineers and scientists. I was very fortunate to interact and have as semiconductor mentors and guidance from my first years at IBM: Ron Troutman, Wendell Noble, Badih El-Kareh, Ed Nowak, Eric Adler, George Sai-Halasz, Tak Ning, Denny Tang, Robert Dennard, Matt Wordeman, Jack Y.-C. Sun, John Aitkens, Paul Bakeman, Andre LeBlanc, Andre Forcier, Bruno Aimi, Charles Stapper, Peter Cottrell and Dale Critchlow. In semiconductor development, I would also like to thank peers and co-workers Michael Hargrove, James Slinkman, Stephen Furkay, Jeffery B. Johnson, Stephen Geissler, Toshi Furukawa, Jim Nakos, Jim Adkisson, Terry Hook, and Chris Long. I would like to thank my managers John Hiltebeitel, Jim McNichol, Mark Hakey, Jim Dunn and Stephen St. Onge in the support of my ESD learning. I would like to thank the IBM failure analysis team from James Never, Peter Czahor, and all the others for years of diagnosing ESD failures to assist in the ESD learning process.

Additionally, in ESD design and development, I would like to thank the circuit designers and I/O teams where together we learned the interaction and issues of circuits and technology for the last 22 years: Roy Flaker, Jack Gersbach, Russel Houghton, Jeffery Chu, Richard Parent, Howard Kalter, H.S. Lee, David Pricer, Thomas Maffit, Jeffery Dreibelbis, Charles Drake, David Hui, Daniel Dreps, Robert Williams, Daniel Young, John Bialas, Roger Gregor, Harold Pilo, Geordie Braceras, Tony Correale, Douglas Stout, Ron Piro, Jeffery Sloan, James Pequignot, Francis Chan, Sam Ray, and others. Outside of IBM, I am indebted to circuit designers from the microprocessor teams of Motorola/IBM, AMD, Digital, Compaq, NEXGEN, TransMeta, IDQ, and Cyrix in development of circuits and ESD devices. Special thanks to Gianfranco Gerosa of Motorola, Donald Draper of NEXGEN, and Mark Johnson from Transmeta.

In the ESD discipline, I would like to express my thanks for the years of support from SEMATECH ESD group, the EOS/ESD Symposium, ESD Assocation Device Testing Standards Committees, as well as discussions on ESD protection and latchup: Charvaka Duvvury, Timothy-Maloney, Professor Elyse Rosenbaum, Professor Ming-Dou Ker, Patrick Juliano, Ajith Amerasekera, Professor Fred Kuper, Jeremy Smith, Eugene Worley, Robert Ashton, Warren Anderson, Michael Chaine, Leo G. Henry, Mark Kelly, Jon Barth, David Bennett, Gary Weiss, Corrine Richier, Marise BaFluer, Richard Ida, Natarajan Mahadeva Iyer, M.K. Radhakrishnan, Gianluca Boselli, Brenda McCaffrey, Ann (Fletcher) Rand, Ron Gibson and John Kinnear. I would like to thank the ESD Association office for the support in the area of publications, standards developments, and conference activities, with a special thanks to Lisa Pimpinella. I would like to thank the International Reliability Physics Symposium staff, and committees for the opportunities to provide ESD lectures and tutorials on ESD and latchup. I would like to acknowledge support from the International Physical and Failure Analysis Symposium, IEEE Reliabiltity Chapter and Chartered Semiconductor in Singapore for the ESD support in developing ESD tutorials. A special thanks to Professor Ming-Dou Ker, and Professor Steve Chung of National Chiao-Tung University and the Taiwan ESD Symposium. A special thanks to Professor Shun-Lien Chuang and Professor Elyse Rosenbaum of University of Illinois Champaign-Urbana. I would like to thank my summer students Kimberly Morriseau of Essex Junction Vermont, Patrick Juliano of UIUC, Brian Ronan of Princeton University, and Anne Watson of Penn State University for hard work, collaboration, and participating in ESD and latchup experimentation, and discovery.

And most important to my children, Aaron Samuel Voldman, and Rachel Pesha Voldman, and my wife Annie Brown Voldman, for supporting and putting up with this effort, and for keeping our lives going forward, with all the demands of life and desires that we have, and to make the sacrifices necessary to keep our lives on the Right Path. And of course, my parents, Carl and Blossom Voldman and family.

B''H
Steven H. Voldman
IEEE Fellow

1 Electrostatics and Electrothermal Physics

1.1 INTRODUCTION

The discovery of electrostatic attraction and electrostatic discharge (ESD) is one of the earliest examples of the understanding of scientific thought and analysis. It goes back to the early foundations of the problem of the nature of matter, astronomy, mathematics and philosophy, and predates understanding of the nature of matter.

Thales of Miletus (624–546 B.C.E) was the founder of the Ionian (or Milesian) school and one of the Seven Wise Men of Ancient Greece in the pre-Socratic era. Thales was an astronomer, mathematician, and philosopher. He was an inventor and an engineer. Thales established a heritage of searching for knowledge for knowledge's sake, development of the scientific method, establishment of practical methods, and application of a conjectural approach to questions of natural phenomena. The Milesian School is regarded as establishing the critical method of philosophy of questioning, debate, explanation, justification and criticism. Students of Thales included Euclid, Pythagoras and Eudemus [1].

It was Thales of Miletus who was credited with the discovery of the electrostatic attraction. Thales noted that, after amber was rubbed, straw was attracted to the piece of amber. It was from this event that, the Greek word for amber, ελεκτρον (*electron*) became associated with electrical phenomena. Knowledge of Thales' ideas was common through the writings of his disciples and Greek philosophers. Aristotle stated 'Some think that the soul pervades the whole universe, whence perhaps came Thales' view that everything is full of gods.' [2].

Investigation of electrostatic phenomena predated early thoughts on the nature of physical matter. Thales's ESD experiments and study of electrostatic attraction occurred before the Greek atomistic schools of Democritus (420 B.C.E.), and Epicurus (370 B.C.E.), and the Roman school of Lucretius (50 B.C.E). Thales died before the schools of atomic thought were active. His tomb was inscribed 'Here in a narrow tomb Great Thales lies; Yet his renown for wisdom reached the skies [3].'

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

Robert A. Millikan, in the Introduction of his 1917 edition of *The Electron* [4] stated:

> *Perhaps it is merely a coincidence that the man who first noticed rubbing of amber would induce in it a new and remarkable state now known as the state of electrification was also the man who first gave expression to the conviction that there must be some great unifying principle which links together all phenomena and is capable of making them rationally intelligible; that behind all the apparent variety and change of things there is some primordial element, out of which all things are made and the search for which must be the ultimate aim of all natural science. Yet if this be merely coincidence, at any rate to Thales of Miletus must belong a double honor. For he first correctly conceived and correctly stated, as far back as 600 B.C., the spirit which has actually guided the development of physics in all ages, and he also first described, though in a crude and imperfect way, the very phenomenon the study of which has already linked together several of the erstwhile isolated departments of physics, such as radiant heat, light, magnetism, and electricity, and has very recently brought us nearer to the primordial element than we have ever been before.*

J. H. Jeans, in the 1925 fifth edition of *The Mathematical Theory of Electricity and Magnetism* [5], wrote: 'The fact that a piece of amber, on being rubbed, attracted to itself other small bodies, was known to the Greeks, the discovery of this fact being attributed to Thales of Miletus.' This fact, and the discovery of magnetism from lodestone, formed the 'basis from which the modern science of Electromagnetism has grown.'

With the death of Thales of Miletus, little progress was made on ESD phenomena. Although history moved forward, the advancement of triboelectric charging and electrostatic discharge phenomenon was slow. Europe saw the Roman Empire, the Golden Age of Islam, the Middle Ages, the Black Death, the Renaissance, the Reformation, and the advancement of nation states. ESD was discovered by Thales while China was under the Zhou Dynasty. Asia experienced tremendous changes, through the Qin, Han, Sui, Tang, and Song, Yuan, and Ming dynastic periods, but with no advancement of this field of knowledge.

With all this social change, there was insignificant growth in the understanding of electrostatic phenomenon until the 18th century. Interest in tribocharging and electrostatic phenomena became a hobby of scientists supported by the courts of Europe and laboratories of France and England. Gilbert, in the 17th century, noted that interaction between a glass rod and silk produced the same phenomenon discussed by Thales of Miletus; materials when rubbed with silk became 'amberized' [4]. Gilbert began construction of the earliest list of triboelectrification.

In the same period, Stephen Gray (1696–1736) thought of the concept of dividing of materials according to their nature of removing or sustaining electrification [4]. He defined a class of materials which remove 'amberization' as conductors, and the class of materials which allowed a body to retain its electrification as nonconductors, or insulators.

This work was followed by French physicist Dufay in 1733 discovering that the same effect can be achieved with sealing wax and cat's fur, and noting the effect was different from that of a glass rod [4]. Dufay first noted that there was an attractive and repulsive phenomenon between different materials, naming the opposite processes as 'vitreous' and 'resinous.'

Benjamin Franklin, the first American ESD engineer, in 1747, also identified two processes, which he divided into 'positive' and 'negative' processes. The 'positive' process was the first process, discovered by Gilbert, that a physical body was electrified positive if repelled by a glass rod which had been rubbed with silk. The 'negative' process is any body repelled from sealing wax which had been rubbed with cat's fur, extending the work of Dufay.

In this time frame, many electrostatic scientists began recording the relationship of one body to another in the electrification process. Lists of materials were made to construct the early triboelectrification chart [5]:

Cat's skin, Glass, Ivory, Silk, Rock Crystal, The Hand, Wood, Sulphur, Flannel, Cotton, Shellac, Caoutchouc, Resins, Guttapercha, Metals, and Guncotton

Around this time, some electrostatic scientists and engineers explored phenomena around them, investigating ESD phenomena of garments, and possibly influencing the 'human body model'. Robert Symmer (1759) conducted experimental studies in the dark, exploring the electrical discharge phenomenon on removal of his stockings and the interactions of two different sets of stockings, placing two white stockings in one hand and two black stockings in the other [4].

In contrast, Coulomb developed the torsion balance in 1785, beginning the serious process of the understanding of the relationship of force, charge and physical distance. The experiments of Coulomb had been performed by Cavendish at an earlier date, but not published until 1879 by Maxwell [6].

At this time, the relationship of positive and negative electrification were not fully understood. In 1837, Michael Faraday performed the 'ice-pail experiment' involving a glass rod and silk which showed that 'positive and negative electrical charges always appear simultaneously and in exactly equal amounts.' Faraday, in his lecture *Forces of Matter: Lecture V Magnetism–Electricity* provided demonstrations of electrostatic charging phenomena, demonstrating the relationship of positive and negative charging effects [7]. He would close his lecture on electrical phenomena with 'This, then is sufficient, in the outset to give you an idea of the nature of the force which we call ELECTRICITY. There is no end to the things from which you can evolve this power.'

Even at this time frame, the understanding of the relationship of matter, electrical charging and electrical force were not well understood. Different models were proposed, including single electrical fluid models, and two-fluid models, to explain the charging process. Models were established to explain these phenomena as related to a stress or strain of the medium, moving from an atomistic perspective to a field representation. It was at this time frame that Faraday and James Clerk Maxwell began addressing the understanding of electricity and electrical forces in terms of a field perspective and electrical charge was viewed as a "state of strain in the ether." James Clerk Maxwell, in *Electricity and Magnetism* in 1873, created the formulation of electricity and magnetism as we understand it today [6].

In 1889, Paschen began analysis of the breakdown phenomenon in gases, trying to explain the relationship of breakdown in gases by relating gas pressure and electrode spacing [8]. Breakdown phenomenon in media took a great leap forward, influencing ESD understanding in today's devices.

In 1891, the word 'electron' as a natural unit of electricity was suggested by Dr G. Johnstone Stoney, connecting the ideas of Faraday's Law of Electrolysis [4]. At

this time, the understanding of the connection to physical matter was not understood, but it was used as a measure or unit. Johnstone Stoney reconnected the idea to the Greek word for amber, relating the early work of Thales of Miletus to the modern concept of an electrical unit, later to be shown to be connected to the structure matter and atomic theory.

From this brief history, it can be seen that electrostatic attraction predates the earliest thoughts on the understanding of matter. From 600 B.C.E to even as late as the 1890s the relationship of the discovery of Thales of Miletus were not understood as connected to the transfer of electrons.

Modern physics opened the door to the understanding of the atom and the relationship to electrical charge. It is too vast an undertaking to begin this book with a discussion of modern physics and the evolution of the electrical engineering discipline.

The electrical engineering profession has been compartmentalized into different fields, disciplines and sub-disciplines. It was the microelectronic industry that has led to a return of interest in electrostatic discharge phenomena. In the last 25 years, with the growth of the semiconductor industry, ESD has been a discipline of significant interest. In the 1970s as a result of the concerns in the Cold War of electromagnetic pulses (EMP) excellent physical models were being produced to understand the ESD robustness of electrical components. In the 1980s with the growth of semiconductor fabrication and the semiconductor industry, ESD issues remained an issue as human handling, tooling, shipping and garments, influenced the reliability of semiconductor components. Although ESD concerns existed, the resources and growth was limited in the 1980s, leading to a primitive understanding of how to avoid failure mechanisms, and how to design ESD robust semiconductor processes, devices and circuits. In the 1990s, there has been tremendous growth in ESD understanding, but new fields and new issues continue to emerge. As we make the transition from microelectronics to nanostructures, all types of components, from semiconductors, magnetic recording devices, micromachines, to photomasks have a larger concern with ESD.

Electrostatic discharge is a cross-discipline, bridging thermal, mechanical, and electrical physics. A difficulty in its teaching is that it also spans physical scales from the microscopic to macroscopic. The discipline involves understanding from the atomic surfaces, to nanostructure devices, to circuits, and systems. It spans both space and time. In time, it spans a wide timescale, from picosecond phenomenon to steady state. This is the modern difficulty of understanding of quantification and prediction as it spans disciplines, physical scales, temporal scales, and modeling representation. To conquer this, we will rebuild our understanding and the focus will be on physical models which will progress to practical examples.

1.2 A TIME CONSTANT APPROACH

1.2.1 ESD Time Constants

To understand physical phenomena, and particularly ESD phenomenon, it is necessary to quantify the scale in both space and time. ESD phenomena involve microscopic to macroscopic scales. ESD phenomena involve electrical and thermal transport on the

scale of nanometers, circuits and electronics on the scale of micrometers, semiconductor chip designs on the scale of millimeters, and systems on the scale of meters. The time scales of interest range from picoseconds to microseconds. Electrical currents of interest range from milliamps to tens of amperes. Voltages range of interest vary from volts to kilovolts. Temperatures vary from room temperature to melting temperatures of thousands of degrees Kelvin. It is the vast ranges of time, space, currents, voltages, and temperature as well as its transition from the microscopic to the macroscopic which makes ESD difficult to model, simulate, and quantify.

To comprehend ESD phenomena, and establish validity of analytical developments, it is important to be able understand what phenomenon is important. By analyzing the physical equations from a time constant approach, equations and understanding can be made rigorous, as well as improving logical clarity. In such a fashion, electrical and thermal phenomena can be understood. By familiarity with the important time constants of interest, our understanding as well as insight will be better served. It is through this process that one can establish a higher intuition in this complex field.

1.2.1.1 ESD events

To understand the role of ESD events and the physical environments, it is important to quantify the characteristic times of an ESD event. ESD events are represented as circuit equivalent models.

1.2.1.2 Human body model characteristic time

A fundamental model used in the ESD industry is known as the human body model (HBM) pulse [9]. The model was intended to represent the interaction the electrical discharge from a human being, who is charged, with a component, or object. The model (Figure 1.1) assumes that the human being is the initial condition. The charged source then touches a component or object through a finger. The physical contact between the charged human being and the component or object allows for current transfer between the human being and the object. A characteristic time of the human body model is associated with the electrical components used to emulate the human being. In the HBM standard, the circuit component to simulate the charged human being is a 100 pF capacitor in series with a 1500 Ω resistor. This network has a characteristic rise time and decay time. The characteristic decay time is associated with the time of the network

$$\tau_{\mathrm{HBM}} = R_{\mathrm{HBM}} C_{\mathrm{HBM}}$$

where R_{HBM} is the value of the series resistor and C_{HBM} is the charged capacitor. This is a characteristic time of the charged source. The HBM characteristic time constant is physically interesting since the time of the pulse is of the order of the thermal diffusion time of many materials used in the semiconductor industry.

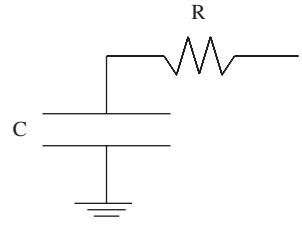

Figure 1.1 Human body model equivalent circuit

1.2.1.3 Machine model characteristic time

Another fundamental model used in the ESD industry is known as the machine model (MM) pulse [10]. The model was intended to represent the interaction the electrical discharge from a conductive source, which is charged, with a component, or object. The model assumes that the 'machine' is charged as the initial condition. The charged source then touches a component or object. In this model (Figure 1.2), an arc discharge is assumed to occur between the source and the component or object, allowing for current transfer between the charged object and the component or object. A characteristic time of the machine model is associated with the electrical components used to emulate the discharge process. In the MM standard, the circuit component is a 200 pF capacitor with no resistive component. An arc discharge fundamentally has a resistance of the order of 10–25 Ω. The characteristic decay time is associated with the time of the network

$$\tau_{\mathrm{MM}} = R_{\mathrm{MM}} C_{\mathrm{MM}}$$

where R is the arc discharge resistor and C is the charged capacitor. This is a characteristic time of the charged source. The MM characteristic time scale is significantly faster than the HBM characteristic time scale, due to the lack of a resistive element. The MM response is oscillatory and has significantly higher currents than the HBM ESD event. Experimentally, MM ESD protection level magnitudes are typically 5–10 times lower than HBM ESD protection level magnitudes.

1.2.1.4 Charged device model characteristic time

The charged device model (CDM) represents an electrostatic discharge interaction between a chip and a discharging means where the chip is precharged [11]. The charging

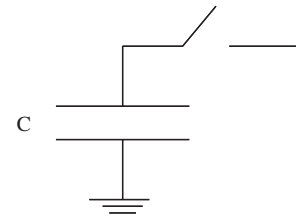

Figure 1.2 Machine model equivalent circuit

process can be initiated by direct charging, or field-induced charging. The discharge process is initiated as contact is initiated between the charged device and the discharging means. CDM discharge occurs at less than 5 ns where typically the rise time of the event is of the order of 250 ps. The CDM event is the fastest of the ESD phenomena, and occurs on a time scale significantly faster than thermal diffusion properties, and of the order of the modern electronic circuit response times. As a result, the response is near the thermal adiabatic assumption for the materials used in semiconductors.

1.2.1.5 Charged cable model characteristic time

Another fundamental model used in the ESD industry is known as the charged cable model or cable model pulse. The model was intended to represent the interaction the electrical discharge of a charged cable, discharging to a chip, card or system. To initiate the charging process, a transmission line or cable is dragged on the floor, leading to triboelectric charging. The model assumes that the cable is charged as the initial condition. The charged cable source then touches a component or object. A characteristic time of the cable model is associated with the electrical components used to emulate the discharge process. In the charged cable model, the cable acts as a capacitor element. The characteristic decay time is associated with the time of the network

$$\tau_{\mathrm{CCM}} = R_{\mathrm{CCM}} C_{\mathrm{CCM}}$$

where R is the discharge resistor and C is the charged cable. The capacitance used for this model is 1000 pF. Studies are also completed using a ESD gun which is discharged through a cable to evaluate the system level events.

1.2.1.6 Charged cassette model characteristic time

The charged cassette model is a recent model associated with consumer electronics. In consumer electronics there are many applications where a human plugs a small cartridge or cassette into a electronic socket. These are evident in popular electronic games. In today's electronic world, there are many palm size electronic components which must be socketed into a system for nonwireless applications. To verify the electronic safety of such equipment, the cassette itself is assumed as a charged source. The cassette model assumes a small capacitance, and negligible resistance. This model is equivalent to a machine model type current source with a much lower capacitor component. The model assumes the resistance of an arc discharge, and a capacitance of 10 pF.

1.2.1.7 Transmission line pulse (TLP) model characteristic time

Transmission line pulse (TLP) testing (Figure 1.3) has seen considerable growth in the electrostatic discharge (ESD) discipline [12]. In this form of ESD testing, a transmission line cable is charged by a voltage source. The transmission line pulse (TLP) system

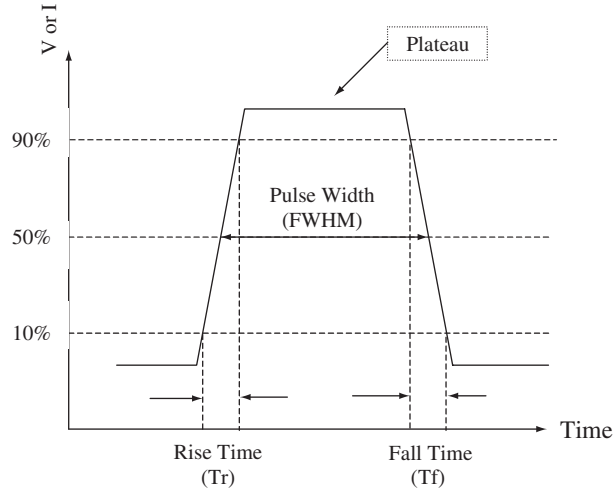

Figure 1.3 Transmission line pulse voltage and current waveform

discharges the pulse into the device under test (DUT). The characteristic time of the pulse is associated with the length of the cable.

The pulse width of a transmission line pulse is a function of the length of the transmission line and the propagation velocity of the transmission line. The propagation velocity can be expressed relative to the speed of light, as a function of the effective permittivity and permeability of the transmission line source.

$$\tau_{\mathrm{TLP}} = \frac{2L_{\mathrm{TLP}}}{\nu} = \frac{2L_{\mathrm{TLP}}\sqrt{\mu_{\mathrm{eff}}\varepsilon_{\mathrm{eff}}}}{c_0}$$

For this method, the choice of pulse width is determined by the decision to use TLP testing as an equivalent or substitute method to the HBM methodology. In standard practice today, the TLP cable length is chosen so as to provide a TLP pulse width of 100 ns with less than 10 ns rise time.

TLP systems are designed in different configurations. TLP system configurations include current source, time domain reflectometry (TDR), time domain transmission (TDT), and time domain reflectometry and transmission (TDRT). In all configurations, the source is a transmission line whose characteristic time constant is determined by the length of the transmission line cable. The various TLP configurations influence the system characteristic impedance, the location of the device under test, and the measurement of the transmitted or reflected signals.

Figure 1.4 is an example of a TDR TLP system. In a TDR TLP system, the pulse is applied to the device, and the evaluation of the waveform is based on the reflected signal.

Figure 1.5 is an example of a time domain transmission (TDT) TLP system. Note that the evaluation of the current and voltage across the device is based on the transmitted signal.

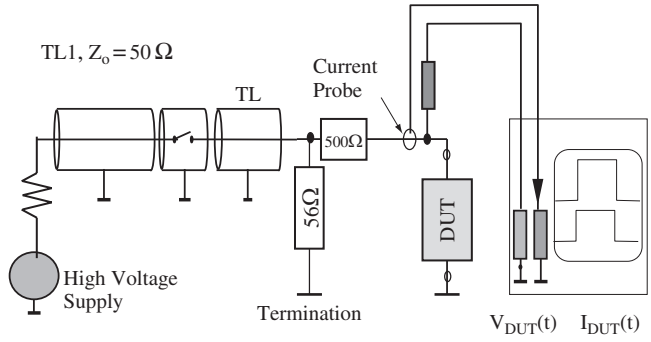

Figure 1.4 Time domain reflection (TDR) transmission line pulse test system

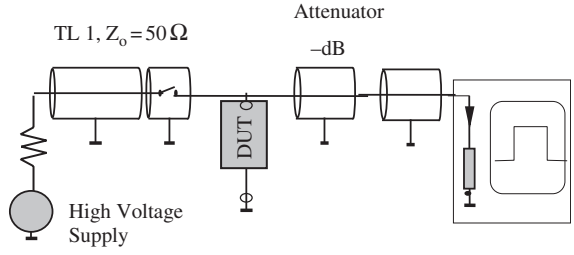

Figure 1.5 Time domain transmission (TDT) transmission line pulse test system

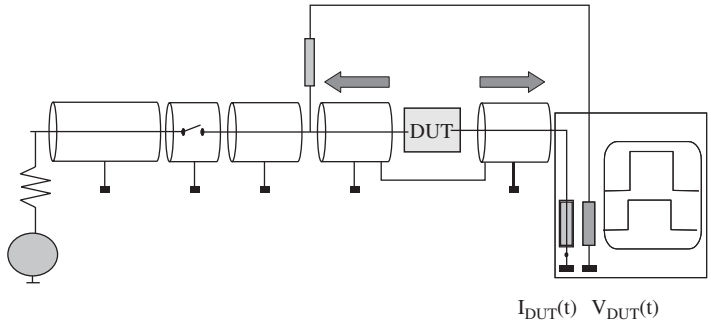

Figure 1.6 Time domain reflection and transmission (TDRT) transmission line pulse test system

In the time domain reflection and transmission (TDRT) TLP system (Figure 1.6), the signal across the device is evaluated based on information of the reflected as well as the transmitted signal.

1.2.1.8 Very fast transmission line pulse (VF-TLP) model characteristic time

The very fast TLP (VF-TLP) test method is similar to the TLP methodology [13]. The interest in VF-TLP is driven by understanding semiconductor devices in a time regime

similar to the charged device model (CDM) time constant. The characteristic time of interest is again determined by the propagation characteristics of the transmission line cable source, and the length of the transmission line cable.

$$\tau_{\text{VF-TLP}} = \frac{2L_{\text{VF-TLP}}}{\nu} = \frac{2L_{\text{VF-TLP}}\sqrt{\mu_{\text{eff}}\varepsilon_{\text{eff}}}}{c_0}$$

The VF-TLP pulse width of interest is a pulse width of less than 5 ns and a rise time < 1 ns. This time regime is well below the thermal diffusion time constant in semiconductor media. The method of the fast time constant limits the suitable equipment for measurement. Presently, VF-TLP systems have been demonstrated using all the same configurations as the TLP system.

1.2.2 Time Constant Hierarchy

By establishing a hierarchy of characteristic times, the validity of the physical phenomena and the equations will be self-evident and an understanding of the time regime under which these will remain valid is obtained [14].

In ESD phenomena, the time constant hierarchy is shown in Figure 1.7 for different ESD mechanisms and simulated tests. The relationship between the pulse time constant and the physical phenomenon will determine the role of thermal and electrical response.

In electrothermal physics, it is important to understand the relationship of the ESD events to the electrical and the thermal response. This is best illustrated by construction of the hierarchy of characteristic length and time scales.

1.2.3 Thermal Time Constant

1.2.3.1 Heat capacity

The heat capacity of a medium is the ability of a medium to store energy [15]. The heat capacity C_{p} of the system is expressed as the product of the mass density ρ, the specific heat c_{p}, and temperature T

$$C_{\text{p}}\Delta T = \rho c_{\text{p}}\Delta T$$

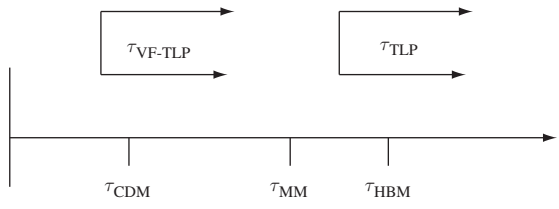

Figure 1.7 ESD time constant hierarchy

1.2.4 Thermal Diffusion

Thermal transport occurs as a result of a gradient in the temperature field. Thermal conduction occurs as a result of thermal diffusion.

1.2.4.1 Heat transport equation

Electrostatic discharge phenomena involve both electrical and thermal phenomena. The temperature field in a medium can be determined from the differential equation of heat conduction. The temperature at any point in the medium can be quantified by understanding the energy balance in a given region. The energy balance equation for an infinitesimal volume is determined by the sum of the net rate of heat entering the volume, and the rate of energy generation in the volume, which is equal to the rate of increase of the internal energy in the volume. The net rate of heat entering the infinitesimal volume is equal to the heat flowing into the volume minus the heat flowing out of the volume. The sum of the differential heat flow in all directions determines the net rate of heating in the volume. This term is the divergence of the heat flux. The rate of energy generation in the volume is associated with the generation sources in the infinitesimal volume. The rate of increase in the internal energy is associated with the increase in the heat capacity of the system. The energy balance equation in this form is also known as the heat equation, or the partial differential equation of heat conduction.

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)+g=\rho c_{\mathrm{p}}\frac{\partial T}{\partial t}$$

In the case that the thermal conductivity k is position and temperature independent the thermal conductivity variable can be separated from the heat flux term. In this case, the partial differential equation of heat conduction can be normalized. The heat flux term can be simplified as the Laplacian of temperature and expressed as

$$\nabla^2 T+\left(\frac{g}{k}\right)=\frac{1}{\alpha}\frac{\partial T}{\partial t}$$

In this form, the thermal diffusivity is defined as

$$\alpha=\frac{k}{\rho c_{\mathrm{p}}}$$

This equation in this form is important for quantifying the ESD event in the location where the heat generation is occurring. This is typically in the regions of electrical current flow or electrical current generation. For example, heat is generated in the source, drain and channel of a MOSFET as a result of Joule heating.

Given that there is no internal generation of heat sources inside the medium, the partial differential equation of heat conduction can be simplified to the Fourier equation, also known as the diffusion equation.

$$\nabla^2 T = \frac{1}{\alpha}\frac{\partial T}{\partial t}$$

The Fourier equation is also known as the parabolic equation since it is a differential equation which is first order in time and second order in space. For ESD analysis, this equation is applied in regions where there is no source of thermal generation, yet whose temperature field is being influenced by heat flux or temperature gradients inside the regions or on its boundaries.

For steady state processes, the partial differential equation of heat conduction simplifies to

$$\nabla^2 T + \left(\frac{g}{k}\right) = 0$$

For the case of a uniform thermal conductivity (space and time), the partial differential equation of heat conduction reduces to the Poisson equation with temperature as the field variable. The Poisson equation is valuable for analysis of self-heating processes which are steady state. In the case that there is no internal generation, uniform thermal conductivity and no heat generation, the Possion equation simplifies to the Laplace equation

$$\nabla^2 T = 0$$

1.2.4.2 Thermal physics time constants

From the partial differential equation of heat conduction, a characteristic time associated with thermal diffusion is the thermal diffusion time τ_{T}

$$\tau_{\mathrm{T}} = \frac{\ell^2}{\alpha}$$

where α is the thermal diffusivity, $\alpha = k/\rho c_{\mathrm{p}}$ and l is the characteristic length.

Given an oscillatory steady state thermal excitation which has an angular frequency ω, where the spatial wavelength is much shorter than other physical characteristic lengths, a thermal skin depth can be defined as

$$\delta = \sqrt{\frac{2\alpha}{\omega}}$$

1.2.5 Adiabatic, Thermal Diffusion Time Scale and Steady State

Electrostatic discharge phenomenon in semiconductors concerns itself with events where the time scale of the applied pulse and response of the medium, circuit and systems are of primary interest.

The hierarchy of characteristic times must be established to understand the physical phenomenon and the relationship of the ESD event characteristic times. The characteristic time of the ESD event relative to the electromagnetic wave transit time, the charge relaxation time, the magnetic diffusion time and the thermal diffusion time are important to understand the physical response.

In semiconductors such as silicon, the hierarchy of characteristic times follow the electroquasistatic (EQS) assumption [14]. For an EQS assumption, the magnetic diffusion time is shorter than the electromagnetic transit time. The ordering of characteristic times are the magnetic diffusion time, the electromagnetic transit time, and then the charge relaxation time. Additionally, the time region of interest is such that the electromagnetic transit time is significantly smaller than the characteristic time of the pulse (e.g. $\beta \ll 1$). The ordering of these characteristic times must be true to insure validity of the EQS assumption. To address the issue of thermal transport, there are three regions of interest within this hierarchy of characteristic times. When the characteristic pulse time τ is much shorter than the thermal diffusion time, the thermal transport due to conduction is negligible. This is the 'adiabatic assumption.' As the characteristic pulse time τ is of the same order of magnitude as the thermal diffusion time. This will be referred to as the 'thermal diffusion regime assumption.' As the characteristic pulse time is significantly longer than the thermal diffusion time, the solution approaches a steady state response, known as the 'steady state assumption.'

1.2.6 Electroquasistatics and Magnetoquasistatics

Given a physical system, a characteristic length and a characteristic time can be established to explain a physical system. Let us define a scale of characteristic length l and characteristic dynamical time τ.

Electrical phenomena involve both electrical and magnetic fields. The electric and magnetic fields are coupled through Maxwell's equations. Three important time constants allow us to understand the validity of the electrical phenomena. The three time constants of interest are the charge relaxation time τ_e, the magnetic diffusion time τ_m, and the electromagnetic wave transit time τ_{em}. The electromagnetic wave transit time τ_{em}, is the time it takes for an electromagnetic plane wave to propagate a distance l

$$\tau_{em} = \frac{l}{c}$$

where c is the speed of light in a medium. The speed of light in a medium can be expressed as

$$c = \frac{1}{\sqrt{\mu\varepsilon}}$$

From this we can express the electromagnetic transit time as

$$\tau_{\text{em}} = \ell\sqrt{\mu\varepsilon} = \sqrt{(\mu\sigma\ell^2)\left(\frac{\varepsilon}{\sigma}\right)}$$

Let us define the charge relaxation time τ_{e}

$$\tau_{\text{e}} = \frac{\varepsilon}{\sigma}$$

and magnetic diffusion time τ_{m}

$$\tau_{\text{m}} = \mu\sigma\ell^2$$

The electromagnetic transit time τ_{em}, can then be expressed as

$$\tau_{\text{em}} = \sqrt{\tau_{\text{e}}\tau_{\text{m}}}$$

In this form, the electromagnetic transit time is the arithmetic mean of the magnetic diffusion time and the charge relaxation time.

Let us define an additional parameter β_τ

$$\beta_\tau = \frac{\tau_{\text{em}}}{\tau}$$

where the parameter is the ratio of the electromagnetic transit time to the characteristic time [14].

For Maxwell's equations to reduce to the EQS assumption, the magnetic diffusion time must be less than the charge relaxation time and the characteristic time τ must be much greater than the electromagnetic transit time (Figure 1.8).

For Maxwells' equations to reduce to the magnetoquasistatic (MQS) assumption, the charge relaxation time must be less than the magnetic diffusion time and the characteristic time τ must be much greater than the electromagnetic transit time (Figure 1.9). In most ESD problems, the analysis does not involve the MQS analysis. But, there are some cases, such as magnetic recording media, magnetic memory, and MRAMs, where the MQS assumption is valid.

In the majority of ESD concerns which are of interest in the scope of the text, electrostatics and electroquasistatics are valid. The characteristic time scale of an ESD pulse will be of the order of a nanosecond to tens of nanoseconds, which is significantly longer than the electromagnetic transit time of a semiconductor device, circuit or chip. For a semiconductor, the substrate materials are such that the charge relaxation time is longer than the magnetic diffusion time as a result of the magnetic permittivity values and electrical permittivity. As the analysis addresses electrical interconnects, packaging, and transmission line pulse systems, the characteristic time approaches the electromagnetic transit time. In the analysis of ESD events involving the arc discharge, both current phenomena, and electromagnetic emissions are present of a TE, TM and TEM form. For semiconductors, the critical time constants are the charge relaxation time, the thermal diffusion time and the time constant of the ESD event, and the circuit response (Figure 1.10).

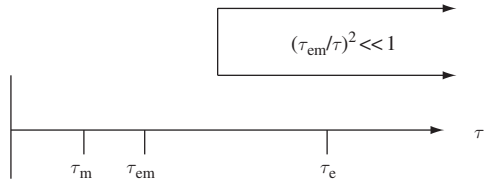

Figure 1.8 Electroquasistatic time constant hierarchy

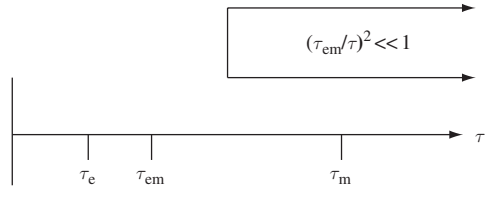

Figure 1.9 Magnetoquasistatic time constant hierarchy

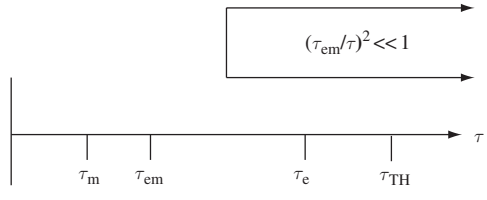

Figure 1.10 Electroquasistatic and thermal time constant hierarchy

1.3 INSTABILITY

1.3.1 Electrical Instability

Electrical instability is associated with a system where the instability involves only electrical phenomena. In this development, it is assumed that the temperature is fixed and has no influence on the stability.

Taking a simple element, such as a resistor, we can state from the relationship of a resistor that

$$V = IR$$

Taking the derivative of this expression in time we obtain,

$$\frac{dV}{dt} = R\frac{dI}{dt} + I\frac{dR}{dt}$$

Assuming voltage is a constant, then $dV/dt = 0$, and

$$0 = R\frac{dI}{dt} + I\frac{dR}{dt}$$

This can be put in the form,

$$0 = \left\{\frac{1}{I}\right\}\frac{\mathrm{d}I}{\mathrm{d}t} + \left\{\frac{1}{R}\right\}\frac{\mathrm{d}R}{\mathrm{d}t}$$

In order to satisfy this relationship, if the first term is positive, then the second term must be negative, or

$$\text{If} \quad \left\{\frac{1}{I}\right\}\frac{\mathrm{d}I}{\mathrm{d}t} \geqslant 0 \quad \text{then} \quad \left\{\frac{1}{R}\right\}\frac{\mathrm{d}R}{\mathrm{d}t} \leqslant 0$$

This expression can be satisfied if $R > 0$, then $\mathrm{d}R/\mathrm{d}t < 0$. This instability condition is such that if the current increases in time, then the resistance must be decreasing in time, or a simpler form, letting $V = IR$ where V is a constant, then

$$R\mathrm{d}I + I\mathrm{d}R = 0$$

Then a condition of instability is

$$\frac{\mathrm{d}R}{\mathrm{d}I} = -\frac{R}{I}$$

Note in this derivation, there was no evaluation of the temperature of the system.

1.3.2 Electrothermal Instability

Electrothermal stability involves the relationship of electrical parameters and temperature [16]. We can establish an instability condition in a general system to define when electrothermal instability can occur. Electrothermal instability is important is the leading cause of ESD failure of devices. Electrothermal instability occurs in resistors, diodes, bipolar transistors and MOSFET devices. As such, we can address the instability criteria independent of the device or structure.

Let current be a function of voltage and temperature, expressing it as

$$I = I(V, T)$$

The total differential of I is then

$$\mathrm{d}I = \left.\frac{\partial I}{\partial V}\right|_T \mathrm{d}V + \left.\frac{\partial I}{\partial T}\right|_T \mathrm{d}T$$

This can be expressed as

$$\mathrm{d}I = \left.\frac{\partial I}{\partial V}\right|_T \mathrm{d}V + \left.\frac{\partial I}{\partial T}\right|_V \frac{\mathrm{d}T}{\mathrm{d}V}\mathrm{d}V$$

From the differential current, $dI = G(V,T)\,dV$, where

$$G(V,T) = \left.\frac{\partial I}{\partial V}\right|_T + \left.\frac{\partial I}{\partial T}\right|_V \frac{dT}{dV}$$

For a stable system, the conductance is a positive, or

$$\left.\frac{\partial I}{\partial V}\right|_T + \left.\frac{\partial I}{\partial T}\right|_V \frac{dT}{dV} > 0$$

but for an unstable system, the conductance is negative, or

$$\left.\frac{\partial I}{\partial V}\right|_T + \left.\frac{\partial I}{\partial T}\right|_V \frac{dT}{dV} < 0$$

When the conductance is positive, and the current is temperature independent for a fixed voltage, such that

$$\left.\frac{\partial I}{\partial T}\right|_V \frac{dT}{dV} = 0$$

and

$$\left.\frac{\partial I}{\partial V}\right|_T > 0$$

then a system is electrically stable.

When the conductance is negative, and the current is temperature independent for a fixed voltage, such that

$$\left.\frac{\partial I}{\partial T}\right|_V \frac{dT}{dV} = 0$$

and

$$\left.\frac{\partial I}{\partial V}\right|_T < 0$$

then a system is electrically unstable.

When the conductance is negative, and the current is temperature independent for a fixed voltage, such that

$$\left.\frac{\partial I}{\partial T}\right|_V \frac{dT}{dV} < 0$$

and

$$\left.\frac{\partial I}{\partial V}\right|_T < 0$$

then a system is electrothermally unstable.

Another way to express this is the temperature where power is the parameter. Let $T = T(P)$, and let $P = VI$, then $T = T(VI)$. Then the partial derivatives of temperature as a function of voltage and temperature can be expressed as

$$\left.\frac{\partial T}{\partial V}\right|_I = I\frac{\mathrm{d}T}{\mathrm{d}P}$$

$$\left.\frac{\partial T}{\partial I}\right|_V = V\frac{\mathrm{d}T}{\mathrm{d}P}$$

The total differential of power can be expressed as

$$\mathrm{d}P = \left.\frac{\partial P}{\partial V}\right|_I \mathrm{d}V + \left.\frac{\partial P}{\partial I}\right|_V \mathrm{d}I$$

$$\mathrm{d}P = \frac{\partial P}{\partial T}\left|\frac{\partial T}{\partial V}\right|_I \mathrm{d}V + \frac{\partial P}{\partial T}\left|\frac{\partial T}{\partial I}\right|_V \mathrm{d}I$$

From $P = VI$,

$$\left.\frac{\partial P}{\partial V}\right|_I = I$$

$$\left.\frac{\partial P}{\partial I}\right|_V = V$$

then

$$\left.\frac{\partial P}{\partial V}\right|_I = \frac{\mathrm{d}P}{\mathrm{d}T}\left.\frac{\partial T}{\partial V}\right|_I = I$$

$$\left.\frac{\partial P}{\partial I}\right|_V = \frac{\mathrm{d}P}{\mathrm{d}T}\left.\frac{\partial T}{\partial I}\right|_V = V$$

Solving for the partial derivatives of temperature as a function of the voltage and current, from the above expressions

$$\left.\frac{\partial T}{\partial V}\right|_I = I\frac{\mathrm{d}T}{\mathrm{d}P}$$

and

$$\left.\frac{\partial T}{\partial I}\right|_V = V\frac{\mathrm{d}T}{\mathrm{d}P}$$

From these expressions we can solve for the differential relation between voltage and current

$$\frac{\mathrm{d}V}{\mathrm{d}I} = \frac{V}{I}\left\{\frac{1 - V\dfrac{\mathrm{d}T}{\mathrm{d}P}\left.\dfrac{\partial I}{\partial T}\right|_V}{R\left.\dfrac{\partial I}{\partial T}\right|_T + V\dfrac{\mathrm{d}T}{\mathrm{d}P}\left.\dfrac{\partial I}{\partial T}\right|_V}\right\}$$

Given that the denominator is nonzero, and $\mathrm{d}V/\mathrm{d}I = 0$ we obtain the condition for second breakdown is

$$1 - V\frac{\mathrm{d}T}{\mathrm{d}P}\left.\frac{\partial I}{\partial T}\right|_V = 0$$

or for current as a function of temperature for a constant voltage,

$$\left.\frac{\partial I}{\partial T}\right|_V = \frac{1}{V}\frac{\mathrm{d}P}{\mathrm{d}T}$$

1.3.3 Spatial Instability and Current Constriction

Instability can occur due to voltage, current or temperature magnitudes but instability can also occur from spatial variations [17]. Electrical and thermal instability can be initiated by drive forces such as voltage gradients or thermal gradients. Thermal instability can be driven by the relationship of temperature as a function of voltage and current, as well as spatial gradients in temperature. Current constriction is the result of a runaway process where self-heating leads to an interaction of the electric field and the thermal field.

Assuming that an equipotential surface is an isothermal (e.g. same temperature). From Ohm's law

$$J = -\left(\frac{1}{\rho}\right)\nabla\phi$$

Where J is the electrical current density and ϕ is the electrical potential field. The electrical resistivity ρ is assumed constant.

The heat flow per unit area can be expressed as

$$q = \phi \bullet J - K\nabla T$$

Assuming conservation of current $\nabla \bullet J = 0$ and conservation of heat $\nabla \bullet q = 0$, by substitution of current density J into the conservation of current relationship, we have

$$\nabla \bullet J = \nabla \bullet \left[-\left(\frac{1}{\rho}\right) \nabla \phi \right] = 0$$

$$\nabla \bullet q = \nabla \bullet \left[\phi \left(-\left(\frac{1}{\rho}\right) \nabla \phi \right) - K \nabla T \right] = 0$$

then

$$\nabla \bullet q = \nabla \bullet [-\phi \nabla \phi - \rho K \nabla T] = 0$$

From the above expression, if the divergence is zero then it must be true that the expression is equal to a constant. Then

$$\phi \nabla \phi + \rho K \nabla T = \text{constant}$$

This expression shows the interrelation of the electric potential and temperature gradient. If the divergence of the heat flux is a constant, as the temperature gradient is increased, this is compensated by decrease in the electric field–potential product expression. At the location of minimum temperature gradient, has the largest potential–electric field product.

At the peak temperature, the gradient of temperature is zero, and the second derivative is minimum. Hence on the contour of peak temperature T_{MAX}, $\bigtriangledown T = 0$. Substituting this into the contour expression, at $T = T_{\text{MAX}}$

$$\phi \nabla \phi = 0$$

A solution to this contour is that the potential is zero. The maximum temperature contour is the minimum potential contour.

$$\int \phi \nabla \phi = - \int \rho K \nabla T$$

where the potential is integrated from potential $\phi = 0$ to potential $\phi = \phi$ and the temperature is integrated from $T_{\max}$ to T.

From the vector identity, where U and V are both scalar fields,

$$\nabla (UV) = U(\nabla V) + V(\nabla U)$$

if we let $U = V = \phi$ then the identity can be expressed as

$$\nabla \left(\frac{\phi^2}{2} \right) = \phi \nabla \phi$$

The integral equation can then be expressed as

$$\int \nabla\left(\frac{\phi^2}{2}\right) = -\int \rho K \nabla T$$

By integral expression the gradient can be interchanged, leaving the expression as

$$\frac{\phi^2}{2} = \int_T^{T_{\max}} \rho(T) K(T) \mathrm{d}T$$

This result shows the electrical energy (square of the potential energy) equals the integral of the product of the resistivity and thermal conductivity integrated from the temperature T to the maximum temperature T_{max}. In this expression, the integration does not follow the isothermal region, but is an integration along the temperature gradient. From this relationship, the electric potential can be evaluated as

$$\phi = \sqrt{2\int_T^{T_{\max}} \rho(T) K(T) \mathrm{d}T}$$

For the temperature range below the intrinsic temperature T_i, the electrical resistivity is determined by the doping concentration n.

$$\rho(T) = \frac{1}{q\mu n}$$

For temperatures between the intrinsic temperature and the maximum temperature, the electrical resistivity can be estimated as a function of the intrinsic carrier concentration

$$\rho(T) = \frac{1}{q\mu n_i}$$

where n_i is the intrinsic carrier concentration. In the case below the intrinsic temperature, the electrical resistivity is a weak function of temperature. For temperatures above the intrinsic temperature, the electrical resistivity is a strong function of the temperature. This is a result of the intrinsic carrier concentration has an exponential dependence in temperature.

From the relationship between electric potential, maximum temperature, electrical resistivity and thermal conductivity, a set of contours (Figure 1.11) can be created as a function of a fixed T_{max}. The maximum temperature in the electrical constriction can be between the intrinsic temperature and the melting temperature.

The radius of the current constriction can be obtained by establishing an Ohm's law relationship from an infinitesimal in the constriction current path. Hence, a differential

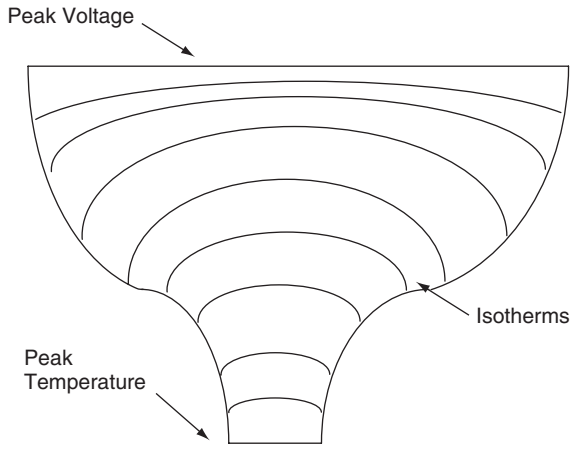

Figure 1.11 Spatial current constriction and temperature contours (isotherms)

change in potential is equal to the product of the total current, the resistivity, and incremental spatial change in the constriction.

$$\mathrm{d}\phi = \mathrm{I_m}\rho\mathrm{d}r$$

The relationship between the differential potential and the differential of thermal potential is

$$\phi\mathrm{d}\phi = \rho K\mathrm{d}T$$

Combining the two equations

$$\mathrm{I_m}\rho\mathrm{d}r = \frac{\rho K\mathrm{d}T}{\phi}$$

Integration over the variable dr, the reduced resistance due to current constriction can be calculated,

$$\int \mathrm{d}r\mathrm{I_m}\rho = \int \mathrm{d}T\frac{\rho K}{\phi}$$

The integration of the geometric dimension is from $r = 0$ to $r = R_\mathrm{s}$, and the integration over temperature is from T_max to T.

Since the total current is a constant, this can be removed from the integral; hence the expression is

$$I_\mathrm{m}R_\mathrm{s} = \int \mathrm{d}rI_\mathrm{m}\rho = \int \mathrm{d}T\frac{\rho K}{\phi}$$

Hence, the reduced resistance can be evaluated as

$$R_s = \left\{\frac{1}{I_m}\right\} \int dT \frac{\rho K}{\phi}$$

From this development, substituting in the expression ϕ (T), we obtain

$$R_s = \left\{\frac{1}{I_m}\right\} \int dT \frac{\rho K}{\sqrt{2\int_T^{T_{msc}} \rho(T')K(T')dT'}}$$

From spreading resistance theory, the relationship between contact radius a and a contour at a distance d from it can be shown to satisfy a model

$$R_s = \left\{\frac{1}{2\pi a}\right\} \tan^{-1}\left\{\frac{d}{a}\right\}$$

Then we have the relationship of the current restriction as

$$\left\{\frac{1}{2\pi a}\right\} \tan^{-1}\left\{\frac{d}{a}\right\} = \left\{\frac{1}{I_m}\right\} \int dT \frac{\rho K}{\sqrt{2\int_T^{T_{msc}} \rho(T')K(T')dT'}}$$

1.4 BREAKDOWN

1.4.1 Paschen's Breakdown Theory

Another form of instability occurs when there is a regenerative feedback. Breakdown in gases, liquids or solids can be initiated by a feedback induced by the acceleration of carriers, leading to secondary carriers.

Breakdown phenomenon in air is important for ESD applications for spark gaps, ESD simulators, and in the magnetic recording industry. Today, there is a focus on the understanding of breakdown phenomenon for charged device simulators. In the magnetic recording industry, breakdown can occur between the magnetoresistor (MR) element and the shields across the air bearing surface. Hence the physics of breakdown in air is relevant to today's problems. At very high speeds, the ability to provide semiconductor devices may be limited. As a result, field emission devices and spark gaps may play a role in air bridge applications and micro-machines.

Paschen, in 1889, studied the breakdown physics of gases in planar gap regions [8,18]. The result showed that breakdown process is a function of the product of the gas pressure and the distance between the electrodes. Paschen showed that

$$pd \simeq \frac{d}{l}$$

where p is the pressure, d is the distance between the plates and l is the mean free path of the electrons. From the work of Paschen, a universal curve was established which followed the same characteristics, independent of the gas in the gap. The Paschen curve is a plot of the logarithm of the breakdown voltage as a function of the logarithm of the product of the pressure and gap distance.

$$V_{\mathrm{BD}} = f(pd)$$

At very low values of the pd product, electrons must accelerate beyond the ionization limit to produce an avalanche process because the probability of impacts is too low. In this region, the breakdown voltage decreases with increasing value of the pressure–gap product. This occurs until a minimum condition is reached. At very high values of the pressure–gap product, the number of inelastic collisions is higher and the breakdown voltage increases. This U-shaped dependence is characteristic of gas phenomenon. At high gas pressure, secondary processes, such as light emissions occur.

1.4.2 Townsend's Concept

The avalanche phenomenon is important to understand the breakdown process in semiconductors and other materials. Townsend, in 1915, noted that breakdown occurs at a critical avalanche height [18,19],

$$H = e^{\alpha d} = \frac{1}{\gamma}$$

In this expression, the avalanche height H is equal to the exponential of the product of the probability coefficient of ionization(number of ionizing impacts per electron and unit distance in the direction of the electric field) and electrode spacing. The avalanche height H can also be expressed as the inverse of the probability coefficient of regeneration (number of new electrons released from the cathode per positive ion).

1.4.3 Toepler's Law

Evaluation of the resistance of arc discharges are important in ESD phenomena since these events are evident in ESD simulation, such as the charged device model (CDM), machine model (MM) and other ESD simulators.

Toepler, in 1906, established a relationship of the arc resistance in a discharge process [18,20]. Toepler's law states that the arc resistance at any time is inversely proportional to the charge which has flowed through the arc

$$R(t) = \frac{k_{\mathrm{T}} D}{\int_0^t I(t')\mathrm{d}t'}$$

where $I(t)$ is the current in the arc discharge at time t, and D is the gap between the electrodes. The parameter k_T is a constant whose value is 4×10^{-5} V s^{-1}cm^{-1}.

1.5 AVALANCHE BREAKDOWN

Avalanche breakdown plays an important role in the understanding of electrostatic discharge issues. Avalanche breakdown can lead to physical failure of semiconductor components. Avalanche breakdown is also used intentionally in semiconductors to serve as a trigger element to initiate turn-on of an ESD circuit. Avalanche breakdown plays a key role in semiconductor ESD elements and circuits which are intentionally reverse biased during functional operation of semiconductor chips, but are initiated at a voltage above the native operational voltage. This phenomenon is key to ESD networks in MOSFET and bipolar devices. It is utilized on input *pin* circuits, between common power rails in ESD power clamps, and between power rails. Hence, it is fundamental to the ESD discipline.

As carriers are accelerated in a medium, energy is transferred from the electric field to the carriers. As the electric field increases, carriers approach a limiting drift velocity and further increases lead to thermal vibrations. As carriers are accelerated, there is competition between energy transmitted to the electron and energy transmitted to the lattice. The probability of ionizing a carrier can be expressed as [21]

$$\exp\left(-\varepsilon_i/qEl_r\right)$$

where the probability is a function of the ionization threshold, the energy of the carrier, and mean free path of optical phonon scattering. Above the threshold for ionization, the energy balance is

$$qE = \alpha_i\varepsilon_i + \alpha_r\varepsilon_r$$

where the LHS is the energy of the carrier, and the RHS is the energy transfer to impact ionization and the energy transfer to optical phonon generation. The impact ionization coefficient is the number of ionizing events divided by distance. From this we can establish a relationship of the probability that impact ionization can occur compared with phonon generation, when the energy of the carrier is above the impact ionization energy.

$$P = \frac{\alpha_i}{\alpha_r}$$

Within that distance, multiple phonons can be generated. This can then be expressed as a function of the impact ionization relative mean free path, and the optical generation mean free path. The probability of emission of an optical phonon is

$$rP = \frac{l_i}{l_r}P$$

In any event, the probability of ionizing is $1/(1+r)$ and the emission of phonons is $r/(1+r)$. From this expression, and substitution into the energy balance equation, we can solve for the impact ionization coefficient

$$\alpha_{\mathrm{i}} = \frac{qEP}{\varepsilon_{\mathrm{r}} + (\varepsilon_{\mathrm{i}} + r\varepsilon_{\mathrm{r}})P}$$

To follow the cascade of carriers, the ionization rate is a function of the summation of probable events of cascades. The probability that a carrier contributes to ionization is a function of a first event where the carrier reaches the ionization and the probability that the first event is ionization. A second term is the probability that the carrier achieves the energy of a phonon and ionization event, and that the first event is phonon emission, followed by a ionization event. This is followed by a third term expressing the probability that an electron achieves the energy of two phonons and an ionization energy, where the probability that two phonons are emitted followed by an ionization event.

$$\begin{aligned} P =& P((E = E_{\mathrm{i}})|\mathrm{ion}) + P((E = E_{\mathrm{i}} + E_{\mathrm{r}})|\mathrm{phonon}|\mathrm{ion}) \\ &+ P((E = E_{\mathrm{i}} + 2E_{\mathrm{r}})|\mathrm{phonon}|\mathrm{phonon}|\mathrm{ion}) + \ldots \end{aligned}$$

Then the number of electrons whose first event is ionization is

$$\frac{1}{1+r}\ \exp\ (-\varepsilon_{\mathrm{i}}/qEl_{\mathrm{r}})$$

The number of electrons whose second event is ionization is

$$\frac{1}{1+r}\left(\frac{r}{1+r}\right)\ \exp\ -\left(\frac{\varepsilon_{\mathrm{r}} + \varepsilon_{\mathrm{i}}}{qEl_{\mathrm{r}}}\right)$$

or equivalently

$$\frac{1}{r}\left(\frac{r}{1+r}\right)^2\ \exp\left(-\frac{\varepsilon_{\mathrm{i}}}{qEl_{\mathrm{r}}}\right)\ \exp\left(-\frac{\varepsilon_{\mathrm{r}}}{qEl_{\mathrm{r}}}\right)$$

Hence it can be shown for the event of two phonons and an ionization

$$\frac{1}{r}\left(\frac{r}{1+r}\right)^3\ \exp\left(-\frac{\varepsilon_{\mathrm{i}}}{qEl_{\mathrm{r}}}\right)\ \exp\left(-\frac{2\varepsilon_{\mathrm{r}}}{qEl_{\mathrm{r}}}\right)$$

From this the general expression can be shown that

$$P = \left(\frac{1}{1+r}\right)\exp - \left(\frac{\varepsilon_i}{qEl_r}\right)\left[1 + \left(\frac{r}{1+r}\right)\exp - \left(\frac{\varepsilon_r}{qEl_r}\right) + \left(\frac{r}{1+r}\right)^2 \exp - \left(\frac{2\varepsilon_r}{qEl_r}\right) + \ldots + \left(\frac{r}{1+r}\right)^n \exp - \left(\frac{n\varepsilon_r}{qEl_r}\right)\right]$$

or

$$P = \left(\frac{1}{1+r}\right)\exp - \left(\frac{\varepsilon_i}{qEl_r}\right)\left[\sum_{n=0}^{n=n}\left(\frac{r}{1+r}\right)^n \exp - \left(\frac{n\varepsilon_r}{qEl_r}\right)\right]$$

where this expression can be placed in the relationship below to solve for the ionization coefficient

$$\alpha_i = \frac{qEP}{\varepsilon_r + (\varepsilon_i + r\varepsilon_r)P}$$

Note in this development, the analysis was treated from a probability and scale length perspective. In this form, this is applicable to other applications where there is a loss of energy to a secondary process.

1.5.1 Breakdown in Air

Analysis of the breakdown in air is valuable for ESD protection to understand ESD simulators, and specifications. For example, Lin and Welsher treated the phenomenon of air discharges to understand the physics of the first charged device model (CDM) test system [22].

Assuming a simple geometry of a gap with a gap spacing d, charge Q, and voltage V, we can evaluate the voltage, electric field, peak current, rise and fall time using the Townsend avalanche relationship. Assume a breakdown voltage and breakdown electric field is achieved in the gap leading to the flow of current across the gap structure. Let the current be a function of the electron drift current term (neglecting diffusion current).

$$I = en_e\nu_d$$

and

$$\nu_d = \mu E$$

We can relate the current to the derivative of the charge Q as a function of time, letting $dQ = (env)dt$, and we can relate the drift voltage to the electric field, and we know

that the electric field E is a function of the voltage and gap spacing $E = V/d = Q/Cd$. Hence we can write

$$\mathrm{d}Q = (en_\mathrm{e}\{\mu E\})\mathrm{d}t = \left(en_\mathrm{e}\left\{\mu\frac{Q}{Cd}\right\}\right)\mathrm{d}t$$

Separating the variables, and integrating,

$$\int \mathrm{d}Q\frac{1}{Q} = \int \mathrm{d}t\left(\left\{\frac{e\mu}{Cd}\right\}n_\mathrm{e}\right)$$

Then the charge can be expressed as

$$Q(t) = Q_0\exp\left\{-\int_0^t \mathrm{d}t'\left(\frac{e\mu}{Cd}\right)n_\mathrm{e}(t')\right\}$$

Assuming a breakdown voltage, and breakdown electric field magnitude, as the initial condition to initiate the discharge, we can also state

$$V(t) = V_0\exp\left\{-\int_0^t \mathrm{d}t'\left(\frac{e\mu}{Cd}\right)n_\mathrm{e}(t')\right\}$$

$$E(t) = E_0\exp\left\{-\int_0^t \mathrm{d}t'\left(\frac{e\mu}{Cd}\right)n_\mathrm{e}(t')\right\}$$

The current is the derivative of the charge, hence

$$I(t) = -\left(\frac{e\mu}{Cd}\right)n_e Q_0\exp\left\{-\int_0^t \mathrm{d}t'\left(\frac{e\mu}{Cd}\right)n_\mathrm{e}(t')\right\}$$

From these expressions, the voltage, electric field, charge and current can be obtained in the gap when the electron density as a function of time is evaluated. Hence, all the terminal information can be expressed as a function of the charge density in time.

From the electron current continuity relationship

$$\frac{\partial n_\mathrm{e}}{\partial t} + \nabla \bullet J_\mathrm{e} = G - R$$

In this form, the generation of electrons is associated with the impact ionization generation. In the assumption that the spatial variation is small compared with the time variation, we can neglect the divergence of current term as an approximation.

Additionally, we can assume electron recombination is negligible. The electron generation rate is the ionization rate times the current flux. Then we can express this as

$$\frac{\partial n_e}{\partial t} = \alpha\{n_e \nu_d\}$$

Solving for the electron population in time

$$n_e(t) = n_0 \exp\left\{ - \int_0^t dt' \alpha \nu_d \right\}$$

The ionization constant is a function of the electric field, and hence the integral expression should be integrated over the electric field instead of time.

From the expression for the electric field,

$$E(t) = E_0 \exp\left\{ - \int_0^t dt' \left(\frac{e\mu}{Cd}\right) n_e(t') \right\}$$

and

$$\frac{d}{dt} E(t) = E_0 \frac{d}{dt} \exp\left\{ - \int_0^t dt' \left(\frac{e\mu}{Cd}\right) n_e(t') \right\}$$

Then

$$\frac{d}{dt} E(t) = E_0 \exp\left\{ - \int_0^t dt' \left(\frac{e\mu}{Cd}\right) n_e(t') \right\} \frac{d}{dt} \left\{ - \int_0^t dt' \left(\frac{e\mu}{Cd}\right) n_e(t') \right\}$$

From Leibniz's rule

$$\frac{d}{dt} E(t) = -E\left\{ \int_0^t dt' \left(\frac{e\mu}{Cd}\right) \frac{d}{dt} n_e(t') + \left(\frac{e\mu}{Cd}\right) n_e(t) \frac{dt}{dt} - \left(\frac{e\mu}{Cd}\right) n_e(0) \frac{d(0)}{dt} \right\}$$

where the first and third term are equal to zero, hence

$$\frac{d}{dt} E(t) = -E\left(\frac{e\mu}{Cd}\right) n_e(t)$$

then

$$dE = -E\left(\frac{e\mu}{Cd}\right) n_e(t) dt$$

Using this representation, the electron density can be solved for as a function of the electric field using the drift relationship

$$\mathrm{d}n_\mathrm{e} = \alpha n_\mathrm{e} \nu_\mathrm{d} \mathrm{d}t = \alpha(\mu E) n_\mathrm{e} \mathrm{d}t$$

Substituting in the differential relationship between electric field and the differential in time, we can put the differential in electron population in the form,

$$\mathrm{d}n_\mathrm{e} = \alpha(\mu E) n_\mathrm{e} \left(-\frac{Cd}{e\mu} \frac{\mathrm{d}E}{E n_\mathrm{e}} \right) = -\left(\frac{Cd}{e} \right) \alpha \mathrm{d}E$$

Solving for the electron density by integration of both sides

$$n_\mathrm{e} = n_0 + \frac{Cd}{e} \int_E^{E_0} \alpha(E') \mathrm{d}E'$$

Then from this expression the current can be solved from the relationship of current and carrier density

$$I(t) = e\mu E \left\{ n_0 + \frac{Cd}{e} \int_E^{E_0} \alpha(E') \mathrm{d}E' \right\}$$

1.5.2 Air Breakdown and Peak Currents

Air breakdown is important in ESD phenomena and the physics of failure. A key parameter of interest is the peak current of the air discharge since a number of ESD failure mechanisms are associated with the peak current magnitude. The current from an air discharge can be expressed as

$$I(t) = e\mu E \left\{ n_0 + \frac{Cd}{e} \int_E^{E_0} \alpha(E') \mathrm{d}E' \right\}$$

The first term is associated with the initial drift current, and can be regarded as small compared with the integral term. Solving for the peak current can be achieved by setting the first derivative with respect to the electric field equal to zero. Taking the derivative of the current with respect to the electric field

$$\frac{\mathrm{d}}{\mathrm{d}E} I(t) = \frac{\mathrm{d}}{\mathrm{d}E} \left\{ \mu E e \frac{Cd}{e} \int_E^{E_0} \alpha(E') \mathrm{d}E' \right\}$$

$$\frac{\mathrm{d}}{\mathrm{d}E}I(t) = \mu Cd\frac{\mathrm{d}}{\mathrm{d}E}\left\{\mu E\int_{E}^{E_0}\alpha(E')\mathrm{d}E'\right\}$$

By differentiation, and using Leibniz' rule

$$\frac{\mathrm{d}}{\mathrm{d}E}I(t) = \mu Cd\left\{\left[\int_{E}^{E_0}\alpha(E')\mathrm{d}E'\right] + E\left[\int_{E}^{E_0}\frac{\mathrm{d}}{\mathrm{d}E}\alpha(E')\mathrm{d}E' + \alpha(Eo)\frac{\mathrm{d}Eo}{\mathrm{d}E} - \alpha(E)\frac{\mathrm{d}E}{\mathrm{d}E}\right]\right\}$$

From differentiation of the integral term, only the third term is nonzero. To find the maximum peak current, the derivative of current is set to zero.

$$\frac{\mathrm{d}}{\mathrm{d}E}I(t) = \mu Cd\left\{\left[\int_{E}^{E_0}\alpha(E')\mathrm{d}E'\right] - E[\alpha(E)]\right\} = 0$$

Then the maximum peak current occurs when

$$E_{\max}[\alpha(E_{\max})] = \left[\int_{E_{\max}}^{E_0}\alpha(E')\mathrm{d}E'\right]$$

Solving for the maximum current

$$I_{\text{peak}} = (\mu Cd)E_{\max}^2\alpha E_{\max}$$

The peak current is then a function of the capacitance, the gap size, the mobility and the ionization at the maximum electric field. From the electron carrier density, substituting in the above current condition

$$n_{\max} = n_0 + \frac{Cd}{e}\int_{E_{\max}}^{E_0}\alpha(E')\mathrm{d}E' = n_0 + \frac{Cd}{e}\{E_{\max}\alpha(E_{\max})\}$$

1.5.3 Air Breakdown and Rise Times

Pulse time constants associated with ESD phenomena are of interest in order to understand the high-current physics, and the response of the circuit to the pulse [22]. There are two situations of interest. First, there is interest in whether the semiconductor device or circuit is responsive to the pulse while it is being stressed. Secondly, for design, circuits such as ESD power clamps are designed to discriminate different ESD pulse waveforms from circuit functional responses. For example, RC-discriminators used in RC-triggered power clamps are used to respond to the ESD pulse waveform, but not to the power up, power down, or d.c. operation.

The pulse rise time of air breakdown is important in evaluation of fast ESD phenomenon. From the relationship of the differential electric field and the differential time

$$\mathrm{d}E = -E\left(\frac{e\mu}{Cd}\right)n_{\mathrm{e}}\mathrm{d}t$$

From the relationship

$$I(t) = en_{\mathrm{e}}v_{\mathrm{d}} = e\mu n_{\mathrm{e}}E$$

the expression for electric field can be solved as a function of time

$$\mathrm{d}E = -\left(\frac{I}{Cd}\right)\mathrm{d}t$$

The current from an air discharge can be expressed as a function of the electric field only

$$I(t) = e\mu E\left\{n_0 + \frac{Cd}{e}\int_E^{E_0}\alpha(E')\mathrm{d}E'\right\}$$

Separating the variables, and then integrating, we can express the relationship as

$$\mathrm{d}t = -\left(\frac{Cd}{I(E)}\right)\mathrm{d}E$$

Then

$$t = \int_0^t \mathrm{d}t' = -\int_{E_0}^{E} Cd\frac{1}{I(E')}\mathrm{d}E' = Cd\int_E^{E_0}\frac{\mathrm{d}E'}{I(E')}$$

Hence the time response is a function of the capacitance, the gap size, and the inverse of the current integrated over the electric field. Assuming the initial current is small compared with the integral term

$$I(E) \simeq e\mu E\left\{\frac{Cd}{e}\int_E^{E_0}\alpha(E')\mathrm{d}E'\right\}$$

Substituting for the current in the response time integral expression

$$t = Cd\int_E^{E_0}\frac{\mathrm{d}E'}{e\mu E'\left\{\left(\frac{Cd}{e}\right)\int_{E'}^{E_0}\alpha(E'')\mathrm{d}E''\right\}}$$

Factoring out the common terms, we can express the rise time as

$$t = \frac{1}{\mu}\left\{\int_E^{E_0} \frac{\mathrm{d}E'}{E'\left\{\int_{E'}^{E_0} \alpha(E'')\mathrm{d}E''\right\}}\right\}$$

In this form the rise time is not a function of the capacitor, or the gap, but just a function of the mobility, and the ionization rate integrated over the electric field. Note that this can be expressed as a function of the drift velocity – rise time product, with the expression on the right associated with a characteristic length associated with the drift velocity – rise time product.

$$\nu_{\mathrm{d}} t = \frac{1}{E}\left\{\int_E^{E_0} \frac{\mathrm{d}E'}{E'\left\{\int_{E'}^{E_0} \alpha(E'')\mathrm{d}E''\right\}}\right\}$$

1.5.4 Mesoplasmas and Microplasmas

The plasma state is regarded as a fourth state of matter where electrons and ions move as two populations in a physical space. In a plasma state, the electron and ion population have distinct physical temperatures. The particle systems are interactive with the electric fields.

Microplasma states are localized plasma states which typically occur below the intrinsic temperature. Microplasma states occur in the negative resistance regimes, and are reversible. These states can occur during electrical instability.

Mesoplasma states are associated with irreversible damage and thermal breakdown, also known as second breakdown. Mesoplasma states are electrothermal instabilities. Mesoplasma states occur above the intrinsic temperature and at current densities 10^2–10^3 larger than microplasma states.

1.5.5 Mesoplasma Phenomena

Mesoplasma phenomena can be evaluated using solutions to the heat diffusion equation, assuming a spherical source. As an example, Tasca assumed that the defect where breakdown occurred was spherical in nature [23]. Assuming a spherical geometrical defect in an infinite medium whose temperature is at ambient temperature T_0 and a spherical volume Δ, radius r and surface area S, the internal energy of the defect is equal to the heat capacity times the temperature. The power to failure is the sum of the transient power stored in the spherical volume Δ, the transient power that diffuses through its surface S, and the steady state power. Tasca derived an expression as

$$P_{\mathrm{f}} = \left\{\frac{\rho C_{\mathrm{p}} \Delta}{t_{\mathrm{f}}} + S\left(\frac{K\rho C_{\mathrm{p}}}{t_{\mathrm{f}}}\right)^{1/2} + \frac{8\pi K r}{3}\right\}(T_{\mathrm{c}} - T_0)$$

At pulse widths significantly less than the thermal diffusion time, the second and third terms are negligible compared with the first term. The first term represents the adiabatic term. For pulse widths of the order of the thermal diffusion time, the second term dominates. As the pulse widths become significantly greater than the thermal diffusion time, the expression is dominated by the third term, which is time independent.

The power to failure model of the spherical defect is very dependent on the geometry of the defect as well as the variables for the volume, surface area and radius. Assuming we can verify that in fact the defect is spherical in nature, and can determine the critical temperature of failure, it is then possible to extract the size of the defect. For very short times, the spherical volume can be predicted as

$$\Delta = \left\{ \frac{P_f t_f}{\rho C_p (T_c - T_0)} \right\}$$

where a plot of volume can be calculated based on the critical temperature to failure, the known heat capacity term, and the input power and pulse width.

To calculate the surface area, a second term can be calculated from the material properties

$$S = \left\{ \frac{P_f (t_f)^{1/2}}{(K\rho C_p)^{1/2} (T_c - T_0)} \right\}$$

By varying the pulse width of the applied pulse, the geometric characteristics of the defect can be determined.

In the next chapter, thermal models and ESD analytical models will be discussed in significant detail. Understanding the response of structures to ESD events is highly dependent on the geometric region of interest, material properties, and the time constants.

PROBLEMS

1.1 Avalanche phenomena are important to understand the breakdown process in semiconductors and other materials. Townsend, in 1915, noted that the breakdown occurs at a critical avalanche height

$$H = e^{\alpha d} = \frac{1}{\gamma}$$

In this expression, the avalanche height H is equal to the exponential of the product of the probability coefficient of ionization (number of ionizing impacts per electron and unit distance in the direction of the electric field) and electrode spacing. The avalanche height H can also be expressed as the inverse of the probability coefficient of regeneration (number of new electrons released

from the cathode per positive ion). Derive a relationship assuming the gap is a semiconductor depletion width.

1.2 In testing a semiconductor chip, 'no connects' can store charge when receiving a human body model pulse. Assuming a source of capacitance C and series resistance R is connected to a capacitor C_{pad}, calculate the charge transferred to the pad capacitor.

1.3 In the example above, assume that a second pad is adjacent to the first pad at a spacing of W. Both first and second pads have a solder ball, where, at a given voltage, breakdown occurs between the first and second pad. Calculate the time and voltage at which breakdown occurs between the first and second pad structure. Assuming the second pad is connected to an ESD network whose impedance is zero, calculate the rise time of the second pad.

REFERENCES

1. *Philologus* 1944; 96: 170–182.
2. Aristotle. *De Anima*. 411; a7–8.
3. G. S. Kirk, J. E. Raven and M. Schofield. *The Pre-Socratic Philosophers*, 2nd edn, Cambridge University Press, 1995.
4. R. A. Millikan. *The Electron*, University of Chicago Press, 1917.
5. J. H. Jeans. *The Mathematical Theory of Electricity and Magnetism*, 5th edn, Cambridge University Press, 1925.
6. J. Clerk Maxwell. *A Treatise on Electricity and Magnetism*, 1873.
7. Michael Faraday. *The Forces of Matter, Lecture V: Electricity, Scientific Papers, Harvard Classics*, vol. 30, pp. 62–74, P. F. Colliers & Sons, New York, 1910.
8. F. Paschen. *Annalen der Physik* 1889; **37**: 69.
9. HBM-ESD Sensitivity Testing: Human Body Model (HBM)-Component Level, ESD Association Standard, S5.1, 1993.
10. MM-ESD Sensitivity Testing: Machine Model (MM)-Component Level, ESD Association Standard S5.2, 1994.
11. CDM-ESD Sensitivity Testing: Charged Device Model (CDM)-Component Level ESD Association Standard, D5.3, 1997.
12. TLP-ESD Sensitivity Testing: Transmission Line Pulse (TLP)-Component Level ESD Association Standard Practice SP 5.5, 2004.
13. VF-TLP-ESD Sensitivity Testing Very Fast Transmission Line Pulse (VF-TLP) Component Level ESD Association Standard Practice, Draft A, September 2003.
14. J. Melcher. *Continuum Electro-mechanics*, MIT Press, 1983.
15. M. Necati Ozisik. *Basic Heat Transfer*, McGraw Hill, 1977.
16. C. Popescu. 'The thermal runaway mechanism of second breakdown phenomenon.' *Solid State Electronics* 1970; **13**: 887–901.
17. S. K. Ghandi, *Semiconductor Power Devices*, Wiley, 1977.
18. A. Von Hippel. 'Conduction and breakdown.' In: *The Molecular Designing of Materials and Devices*, MIT Press, 1965, pp.183–197.
19. J. S. Townsend. *Electricity in Gases*, Clarendon Press, Oxford, 1915.
20. Toepler, *Ann. Phys*, 1906; **4**: 191.
21. J. L. Moll, *Physics of Semiconductors*, McGraw Hill, 1964.

22. D. Lin and T. Welsher. 'From lightning to charged device model electrostatic discharges.' *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1992; 68–75.
23. D. Tasca. 'Pulse Power Modes in Semiconductors,' *IEEE Transactions on Nuclear Science* 1970; **NS-17** (6): 364–372.

2 Electrothermal Methods and ESD Models

2.1 ELECTROTHERMAL METHODS

In this chapter, mathematical methods of physics associated with the thermal diffusion equation are first introduced to develop the base for analysis of electrostatic discharge models. Mathematical and practical techniques are introduced such as Green's functions, method of images, transform pairs, flux-potential transfer matrices, Kirchoff transformation, Boltzmann transformation, and the Duhamel principle [1–3]. This establishes the base of knowledge to fully appreciate the assumptions and methods in the development of the solutions and models of Tasca [4], Wunsch and Bell [5], Smith and Littau [6], Ash [7], V. I. Arkihpov *et al*. [8], Vlasov and Sinkevitch [9] and Dwyer *et al*. [10]. The chapter closes by addressing ESD from a probability perspective, introducing the ESD probability distribution functions and cumulative distribution functions [11–13].

2.1.1 Green's Function and Method of Images

To provide a generalized formulation for prediction of power to failure and the transient temperature, a Green's function approach can be utilized [1–3]. From the time-dependent heat equation, with a constant thermal conductivity and constant specific heat–mass density product, the Laplacian of temperature is proportional to the partial derivative of temperature as a function of time. Assuming that there exists a source of heat in an infinite medium at a point (x',y',z') at time t', we can express the temperature at a point in space at point (x,y,z) at time t. For a point in space (x,y,z) there is a spherical shell of sources which emits heat energy at time t'. The spherical shell can be defined as a radial shell of thickness dr', whose spherical radius is

$$r^2 = (x - x')^2 + (y - y')^2 + (z - z')^2$$

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

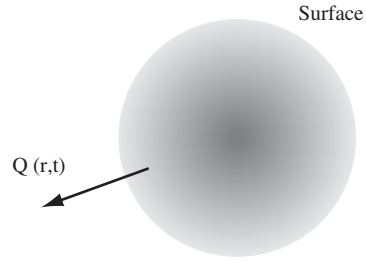

Figure 2.1 Spherical source

A heat source at a point (x′,y′,z′) is generated at time t′ for a quantity of heat $Q\rho c$

$$T(x,x';y,y';z,z';t,t') = \frac{Q}{8[\pi\kappa(t-t')]^{3/2}}\exp\left\{-\frac{r2}{4\kappa(t-t')}\right\}$$

with the spherical radius r (Figure 2.1).

The heat generated at time t' requires a diffusion time associated with the thermal diffusivity to diffuse across the distance between the point the point (x',y',z') to influence the temperature at a time t, at point (x,y,z). This can be generalized by allowing the integration over time using a generalization of the heat at any time t' between the time $t=0$ and the time t.

Hence we can generalize as

$$T(x,x';y,y';z,z';t) = \int_{t'=0}^{t'=t} \mathrm{d}t' \frac{Q(t')}{8[\pi\kappa(t-t')]^{3/2}}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}$$

Expressing this as a function of power dissipated in an infinitesimal volume formed between x' and $x'+\mathrm{d}x'$,y' and $y+\mathrm{d}y'$, and z' and $z+\mathrm{d}z'$, and defining the power dissipated as normalized to the volume V

$$T(x,x';y,y';z,z';t) = \frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t} \mathrm{d}t' \frac{\{P(t')\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'\}}{[(t-t')]^{3/2}}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}$$

To evaluate the temperature over all space from all infinitesimal volumes, the expression can be integrated over the infinite volume Ω. In the expression, the spatially independent terms can be removed from the integration over space, and the temperature can be expressed as

$$T(x,y,z;t) = \frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t} \mathrm{d}t' \frac{P(t')}{[(t-t')]^{3/2}}\int_{\Omega}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$$

2.1.1.1 Parallelepiped in an infinite medium

For evaluation of a parallelepiped in an infinite medium let us assume a source has the dimension W in the x-dimension, and L in the y-dimension, and H in the z-dimension. For electrostatic discharge evaluation, this can represent a structural feature of a physical element. The element can be an interconnect, a resistor element, or a region of a semiconductor MOSFET or bipolar transistor region.

From the above expression, the temperature can be evaluated as

$$T(x,y,z;t) = \frac{1}{8\rho c V(\kappa)^{3/2}} \int_{t'=0}^{t'=t} \frac{P(t')\mathrm{d}t'}{[(t-t')]^{3/2}} F(x-x', y-y', z-z', t-t')$$

with

$$F(x-x', y-y', z-z', t-t') = \int_{-W/2}^{W/2} \exp\left\{-\frac{(x-x')^2}{4\kappa(t-t')}\right\} \frac{\mathrm{d}x'}{\sqrt{\pi}}$$

$$\int_{-L/2}^{L/2} \exp\left\{-\frac{(y-y')^2}{4\kappa(t-t')}\right\} \frac{\mathrm{d}y'}{\sqrt{\pi}}$$

$$\int_{-H/2}^{H/2} \exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\} \frac{\mathrm{d}z'}{\sqrt{\pi}}$$

transformation of variables, we can express the dimension as

$$u = \frac{(x-x')}{\sqrt{4\kappa(t-t')}} \text{ and } \mathrm{d}u = -\frac{\mathrm{d}x'}{\sqrt{4\kappa(t-t')}}$$

Under this transformation, the integral expressions can be represented as error functions in the transformed variables. In a single dimension, we can show

$$\int_{-W/2}^{W/2} \exp\left\{-\frac{(x-x')^2}{4\kappa(t-t')}\right\} \frac{\mathrm{d}x'}{\sqrt{\pi}} = \frac{\sqrt{4\kappa(t-t')}}{2} \left\{\frac{2}{\sqrt{\pi}} \int_0^{u(W/2)} \exp\{-u^2\}\mathrm{d}u \right.$$

$$\left. + \frac{2}{\sqrt{\pi}} \int_{u(-W/2)}^{0} \exp\{-u^2\}\mathrm{d}u \right\}$$

The expression for temperature in an infinite medium can be put in the form

$$T(x,y,z;t) = \frac{1}{8C} \int_{t'=0}^{t'=t} P(t')H(x,y,z;t-t')\mathrm{d}t'$$

where V is the volume $V = LWH$, and letting $C = c\rho V$

where we can write the function containing the spatial dependence as the product of the error functions

$$H(x,y,z;t-t') = \prod_{i=x,y,z}\left[erf\left(\frac{(Lx_i/2)+x_i}{\sqrt{4\kappa(t-t')}}\right)+erf\left(\frac{(Lx_i/2)-x_i}{\sqrt{4\kappa(t-t')}}\right)\right]$$

2.1.1.2 Semi-infinite domain

For the case of the semi-infinite region, we can modify the solution given by the parallelepiped in an infinite region, by defining a boundary condition and applying the method of images (Figure 2.2).

Analogous to the electrostatic problem, defining the temperature T at a point P in an infinite space from a point heat source q at point A, and a point heat sink $-q$ at point A', there exists a locus of points where the temperature is zero. This is the plane which bisects the line AA' at right angles. Hence an 'image' of opposite polarity and strength for an infinite medium exists on the opposite side of the plane. The temperature at the plane formed by the points P is equal to zero, and satisfies the Laplace equation in the semi-infinite half-space, except at the point heat source q at point A.

It is also known that, for any temperature field produced by any number of point heat sources, we can select a constant temperature contour (e.g. isotherm) and replace it with a thermal conductor. The heat sources on either side of the isotherm are the images of the other side.

For the case of the adiabatic boundary condition, we require that the derivative of the temperature is zero at the boundary condition. Assume the parallelepiped is displaced below an infinite plane at $z=0$ by distance D. Let the boundary condition at the $z=0$ plane be an adiabatic boundary condition. This form then assumes that the heat flux for all times $t > 0$ is zero for all points in the plane. Given that we are interested in the temperature field in the semi-infinite space below the plane $z=0$, an image source on the upper half of the infinite space can be used to provide the boundary condition of interest in the semi-infinite lower half-space.

Using a parallelepiped of identical dimensions displaced a distance D above the $z=0$ plane, a plane of symmetry is established that has a solution where all points along the plane $z=0$ have the heat flux at the surface is zero.

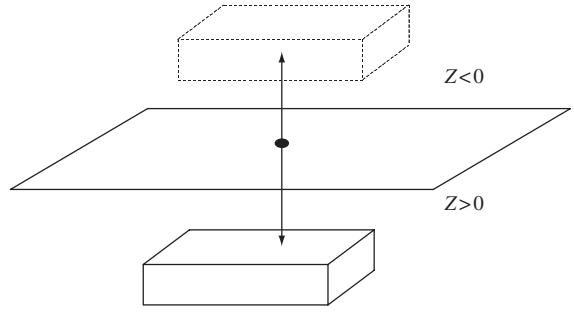

Figure 2.2 Method of images

To solve the problem of the semi-infinite domain, we can utilize the method of images by modification of the Green's function in the infinite domain so that it applies to the semi-infinite domain. From the time-dependent heat equation, with a constant thermal conductivity and constant specific heat–mass density product, the Laplacian of temperature is proportional to the partial derivative of temperature as a function of time. As in the problem of an infinite domain, we have

$$T(x,x';y,y';z,z';t)=\int_{t'=0}^{t'=t}\mathrm{d}t'\,\frac{Q(t')}{8[\pi\kappa(t-t')]^{3/2}}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}$$

Expressing this as a function of power dissipated in an infinitesimal volume formed between x' and $x'+\mathrm{d}x'$,y' and $y+\mathrm{d}y'$, and z' and $z+\mathrm{d}z'$, and defining the power dissipated as normalized to the volume V

$$T(x,x';y,y';z,z';t)=\frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t}\mathrm{d}t'\,\frac{\{P(t')\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'\}}{[(t-t')]^{3/2}}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}$$

To evaluate the temperature over all space from all infinitesimal volumes, the expression can be integrated over the infinite volume Ω. In the expression, the spatially independent terms can be removed from the integration over space, and the temperature can be expressed as

$$T(x,y,z;t)=\frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t}\mathrm{d}t'\,\frac{P(t')}{[(t-t')]^{3/2}}\int_{\Omega}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$$

For evaluation of a parallelepiped in an infinite medium let us assume a source has the dimension W in the x-dimension, L in the y-dimension, and H in the z-dimension, but applying a parallelepiped displaced distance D below the boundary condition $z=0$ and a image source of equal and opposite strength above the $z=0$ plane at $z=\mathrm{D}$

$$T(x,y,z;t)=\frac{1}{8\rho cV(\kappa)^{3/2}}\int_{t'=0}^{t'=t}\frac{P(t')\mathrm{d}t'}{[(t-t')]^{3/2}}F(x-x',y-y',z-z',t-t')$$

with

$$F(x-x',y-y',z-z',t-t')=Fx(x-x':t-t')Fy(y-y';t-t')Fz(z-z';t-t')$$

where

$$F_x(x-x',t-t')=\int_{-W/2}^{W/2}\exp\left\{-\frac{(x-x')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}x'}{\sqrt{\pi}}$$

$$F_y(y-y',t-t')=\int_{-L/2}^{L/2}\exp\left\{-\frac{(y-y')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}y'}{\sqrt{\pi}}$$

$$F_z(z-z',t-t')=\int_{-(D+H)}^{-D}\exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}z'}{\sqrt{\pi}}+\int_{D}^{D+H}\exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}z'}{\sqrt{\pi}}$$

The integral expression can be expressed as error functions using a transformation of variables. The expression for temperature in an infinite medium can be put in the form

$$T(x,y,z;t) = \frac{1}{8C}\int_{t'=0}^{t'=t} P(t')H(x,y,z;t-t')\mathrm{d}t'$$

where $V = LWH$, is the volume and letting $C = c\rho V$ where we can write the function containing the spatial dependence as the product of the error functions

$$H(x,y,z;t-t') = H(z;t-t')\prod_{i=x,y}\left[erf\left(\frac{(L_{\mathrm{x}i}/2)+x_{\mathrm{i}}}{\sqrt{4\kappa(t-t')}}\right)+erf\left(\frac{(L_{\mathrm{x}i}/2)-x_{\mathrm{i}}}{\sqrt{4\kappa(t-t')}}\right)\right]$$

and

$$H(z;t-t') = \left[erf\left(\frac{z+D+H}{\sqrt{4\kappa(t-t')}}\right)+erf\left(\frac{-D-z}{\sqrt{4\kappa(t-t')}}\right)\right.$$
$$\left.+erf\left(\frac{z-D}{\sqrt{4\kappa(t-t')}}\right)+erf\left(\frac{D+H-z}{\sqrt{4\kappa(t-t')}}\right)\right]$$

This solution can be applied to a large plurality of semiconductor and electrical components resistors, transistors, interconnects and magnetic recording devices. In all cases, boundary conditions can be applied to define the parallelepiped element. It is common in ESD analysis that the assumption of the region of interest is based on the geometrical region where electrical current or heat is occurring in the physical elements, insulating boundaries, or associated with a geometrical dimension of the physical element (e.g. junction depth, length or width). The method of images can also be applied to the case of multiple elements, as is commonly done in electrostatics.

2.1.2 Integral Transforms of the Heat Conduction Equation

The heat conduction equation can be solved using integral transforms [1]. The heat conduction equation is of interest only in cases of time greater than or equal to zero on the right-hand side of the space–time axis. Let us define a transform integral pair of temperature using the Laplace transform pair

$$F(x,s) = \int_{t=0}^{t=\infty} e^{-st}T(x,t)\mathrm{d}t$$

and

$$T(x,t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{st}F(x,s)\mathrm{d}s$$

where the function $F(x,s)$ exists only in the right half-plane (R.H.P.) of the s-space where the real part of s satisfies Re $\{s\} > \alpha$. The properties of the function $F(x,s)$ are that it must be analytic in that domain and may be defined by analytic continuation.

From the time dependent heat conduction equation with constant coefficients

$$\frac{\partial^2 T(x,t)}{\partial x^2} = \frac{1}{\kappa}\frac{\partial T(x,t)}{\partial t}$$

the function can be expressed as a function of the transform pair $F(x,s)$. Applying the Laplace transform to both sides of the equation

$$L\left\{\frac{\partial^2 T(x,t)}{\partial x^2}\right\} = L\left\{\frac{1}{\kappa}\frac{\partial T(x,t)}{\partial t}\right\}$$

On the left-hand side of the equation, the Laplace transform of the second derivative of temperature with respect to space can be shown to be equal to the second derivative of the transform variable $F(x,s)$ with respect to space. From Leibniz's theorem, the differentiation and integration can be interchanged, when the boundaries of integration are independent of the differentiation variable. Then

$$L\left\{\frac{\partial^2 T(x,t)}{\partial x^2}\right\} = \int_{t=0}^{t=\infty} e^{-st}\frac{\partial^2}{\partial x^2}T(x,t)\mathrm{d}t = \frac{\partial^2}{\partial x^2}\left\{\int_{t=0}^{t=\infty} e^{-st}T(x,t)\mathrm{d}t\right\} = \frac{\partial^2}{\partial x^2}F(x,s)$$

The right-hand side of the equation can be evaluated using Leibniz's theorem, or the Laplace transform identity for differentiation

$$L\left\{\frac{1}{\kappa}\frac{\partial T(x,t)}{\partial t}\right\} = \frac{1}{\kappa}L\left\{\frac{\partial T(x,t)}{\partial t}\right\} = \frac{1}{\kappa}\left\{\int_{t=0}^{t=\infty} e^{-st}\frac{\partial T(x,t)}{\partial t}\mathrm{d}t\right\}$$

$$= \frac{1}{\kappa}\left\{s\int_{t=0}^{t=\infty} e^{-st}T(x,t)\mathrm{d}t - T(x,0)\right\}$$

For a temporal initial condition that at $t=0$, the temperature at all position is $x=0$, the initial condition term is $T(x,0)=0$. Substituting in the transformed variable

$$L\left\{\frac{1}{\kappa}\frac{\partial T(x,t)}{\partial t}\right\} = \frac{1}{\kappa}\{sF(x,s)\}$$

The time-dependent heat conduction equation in the transform pair $F(x,s)$ is

$$\frac{\partial^2 F(x,s)}{\partial x^2} = \frac{sF(x,s)}{\kappa}$$

To solve the time-dependent heat conduction equation, the boundary conditions must be transformed into the Laplace transform pair function $F(x,s)$. Assuming a boundary condition of the first kind

$$F(x=0,s)=\int_0^\infty e^{-st}T(x=0,t)\mathrm{d}t$$

As a first example, let us treat the problem of a constant step temperature where the temperature is initially zero, and the temperature is raised to a temperature T between the initial temperature and the melting temperature of the boundary medium or the melting temperature within the medium on the left-hand side of the spatial plane ($0 < T < T_M$). The solution of interest will provide the temperature distribution within the medium assuming an instantaneous step increase to a given temperature. In the case where the temperature instantaneously increases to the melting temperature of a first medium, the solution will provide the temperature field in the second medium.

Assuming a boundary condition, $T(x=0,\ t)=T$, using Laplace transformation

$$F(x=0,s)=\int_0^\infty e^{-st}T(x=0,t)\mathrm{d}t=\int_0^\infty e^{-st}T\mathrm{d}t=T\int_0^\infty e^{-st}\mathrm{d}t=\frac{T}{s}$$

In the transformed variable, the partial differential equation is easily solvable compared with the time form as a result of the reduction of the time variable to an algebraic term. Let us assume a solution, for all space $x > 0$

$$F(x,s)=Ae^{\alpha(s)x}+Be^{-\alpha(s)x}$$

As x approaches infinity, we require the function $F(x,\ s)$ tends to zero, hence A must be equal to zero. Solving for boundary condition $F(x=0,\ s)$, the constant A is obtained. Substituting in the transformed time-dependent heat conduction equation, the solution is

$$F(x,s)=\frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\}$$

Using the inverse Laplace transform, we can express the solution for temperature $T(x,t)$ in integral form

$$T(x,t)=\frac{1}{2\pi i}\int_{s=c-i\infty}^{s=c+i\infty}\frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\}e^{st}\mathrm{d}s$$

To evaluate the integral along the contour, a closed contour can be formed. Using a closed contour, the Cauchy integral formula can be applied. Assume a closed contour is formed where a first contour extends from the upper limit of the contour to the LHP s-plane extending to the real axis at a negative infinite value. A second contour extends from a negative infinite real value toward the $s=0$ point. To avoid the singularity, a clockwise circle contour is formed around the singularity. The contour extends from

$s = 0$ to a negative infinity value. This is then followed by a contour that reconnects to the beginning of point above integral expression. From Cauchy integral formula, the closed line integration over a function $f(z)$ where $f(z)$ is analytic in the region and equal to zero.

$$\oint_C f(z)\mathrm{d}z = 0$$

and

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\xi)\mathrm{d}\xi}{\xi - z}$$

Applying the Cauchy theorem for analytical function about a closed contour, the sum of the integral transform, and the integration over the new contour integrated with the function contained in the integral transform is zero. Hence the solution of the temperature is equal to the negative of the sum of the evaluation over the new additional contour terms. Evaluation of the two contours at infinity in the complex plane must be equal to zero. Using the Cauchy residue theorem, the integration around the pole at $s = 0$ is evaluated, where we define the function in the limit that s approaches zero

$$\lim_{s\to 0} f(s) = \lim_{s\to 0} \frac{T\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\}}{2\pi i} e^{st} = \frac{T}{2\pi i}$$

Then evaluation of the pole, equals the value T. For the first contour from negative infinity to zero, the integral is

$$\frac{1}{2\pi i}\int_{s=-\infty}^{s=0} \frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\} e^{st}\mathrm{d}s$$

and from zero to negative infinity below the branch cut, the integral is

$$\frac{1}{2\pi i}\int_{s=0}^{s=-\infty} \frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\} e^{st}\mathrm{d}s$$

By changing variables from s to $-s$, and where the product of value $i^2 = -1$, the variables and boundaries can be transformed, where the boundary is integrated from 0 to infinity.

$$\frac{1}{2\pi i}\int_{s=-\infty}^{s=0} \frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\} e^{st}\mathrm{d}s = -\frac{1}{2\pi i}\int_{s=0}^{s=\infty} \frac{T}{s}\exp\left\{-ix\sqrt{\frac{s}{\kappa}}\right\} e^{-st}\mathrm{d}s$$

and below the branch cut

$$\frac{1}{2\pi i}\int_{s=0}^{s=-\infty}\frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\}e^{st}\mathrm{d}s=\frac{1}{2\pi i}\int_{s=0}^{s=\infty}\frac{T}{s}\exp\left\{ix\sqrt{\frac{s}{\kappa}}\right\}e^{-st}\mathrm{d}s$$

Using the Euler relationship

$$\exp\left\{ix\sqrt{\frac{s}{\kappa}}\right\}=\cos\left\{x\sqrt{\frac{s}{\kappa}}\right\}+i\sin\left\{x\sqrt{\frac{s}{\kappa}}\right\}$$

the two integrals can be combined and expressed as

$$\frac{1}{2\pi i}\int_{s=0}^{s=\infty}\frac{T}{s}\left[\exp\left\{ix\sqrt{\frac{s}{\kappa}}\right\}+\exp\left\{ix\sqrt{\frac{s}{\kappa}}\right\}\right]e^{-st}\mathrm{d}s=\frac{1}{2\pi i}\int_{s=0}^{s=\infty}\frac{T}{s}\left[(2i)\sin\left\{x\sqrt{\frac{s}{\kappa}}\right\}\right]e^{-st}\mathrm{d}s$$

From this expression and the integration around the pole at $s=0$ we can solve for the solution for temperature as a function of the integral

$$T(x,t)=\frac{1}{2\pi i}\int_{s=c-i\infty}^{s=c+i\infty}\frac{T}{s}\exp\left\{-\sqrt{\frac{s}{\kappa}}x\right\}e^{st}\mathrm{d}s=T\left\{1-\frac{1}{\pi}\int_{s=0}^{s=\infty}\frac{\sin\left(x\sqrt{\frac{s}{\kappa}}\right)}{s}e^{-st}\mathrm{d}s\right\}$$

Conversion of the variable s to the square of z allows for the integration of the sine function, where we have

$$T(x,t)=T\left\{1-\frac{1}{\pi}\int_{z=0}^{z=\infty}\frac{\sin\left(z\left[x\sqrt{\frac{1}{\kappa}}\right]\right)}{z}e^{-tz^2}\mathrm{d}z\right\}$$

This integral expression can be expressed as the complementary error function,

$$T(x,t)=T\left[1-erf\left\{\frac{x}{2}\sqrt{\frac{1}{\kappa t}}\right\}\right]=Terfc\left\{\frac{x}{2}\sqrt{\frac{1}{\kappa t}}\right\}$$

From this form, given a constant temperature at the interface, the temperature at any point in the half plane at any time can be expressed as a function of the complementary error function.

From this development, it is clear that the heat equation can be solved in the variable of temperature as a partial differential equation, or the solution can be transformed into an algebraic expression in the transformed variable s.

2.1.3 Flux Potential Transfer Relations Matrix Methodology

In the case of a volumetric region where no heat generation source is present, the Laplace equation is satisfied. In many problems, it is not of interest to know the solution inside the volumetric region, but only of interest to evaluate at the boundaries. Additionally, a virtual boundary can be defined where the solution is of interest at that boundary or interface. The solution of the Laplace equation can be represented as a matrix representation in terms of the relationship between the heat flux and temperature at two interfaces. In this form, the solution can be solved using a 'transfer matrix' (Figure 2.3) where the transfer matrix and its corresponding boundary condition satisfies the Laplace equation. This methodology can be used in planar, cylindrical and spherical geometric regions [14].

In the case of a region with boundaries, or material differences, a stratified medium can be represented as a set of n regions where the solutions at the interfaces are of interest, or only the relationship of the first and last outer boundaries. Multiple films of planar films can be represented as a stratified medium where the material properties may vary in the different physical regions. For the electroquasistatic analysis, the medium can change in the dielectric permittivity and magnetic permittivity. For the thermal analysis, the medium can change in its thermal conductivity. Note that any physical regions with a variable electrical or thermal properties can be segmented into a plurality of regions where an average electrical or thermal property is defined in that region.

This methodology is valuable in many semiconductor problems. For example, in semiconductor chips, there are presently different types of insulator films of different electrical and thermal properties. In the interlevel dielectric (ILD) regions, the insulator region can vary from silicon dioxide to 'low-k' materials which have different thermal conductivities. A second example is in silicon on insulator (SOI) technology; the wafer can be modeled as a silicon film, a insulator film, followed by a silicon substrate. A third case, is that of buried layers such as sub-collector regions or heavily doped buried layers (HDBL). These cases can be modeled as a stratified medium for thermal analysis and prediction of the temperature and heat flux. This methodology is valid for both the thermal and the electrical analysis.

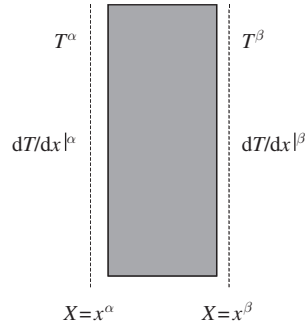

Figure 2.3 Transfer matrix formulation

For a charge-free region with a material of uniform permittivity

$$D = \varepsilon E$$

and

$$P = (\varepsilon - \varepsilon_0)E$$

From Maxwell's equations, the Gauss law relationship is

$$\nabla \bullet D = 0$$

From the relationship of electrical field to the potential field,

$$E = -\nabla\phi$$

and then the solution within the charge free volume satisfies Laplace's equation

$$\nabla^2\phi = 0$$

and

$$D = -\varepsilon\nabla\phi$$

In magnetoquasistatics, analogously given a system with a uniform permeability, a magnetic potential and magnetic flux density can be defined. Since the divergence of the magnetic flux density is zero, from Maxwell's equations, a magnetic potential can satisfy the Laplace's equation, and we can define a relationship

$$B = -\mu\nabla\phi_m$$

To evaluate the thermal flux and the temperature field, we also have the thermal analogy of the system with the uniform thermal properties satisfying the Laplace equation and the flux potential relationship of

$$q = -k\nabla T$$

Addressing the thermal problem, the relationship between the thermal flux from the thermal potential on the boundaries, let us find a matrix representation of the relationship of the forward heat flow from the boundary conditions of temperature. Conversely, if we know the flux on the boundaries, we can determine the potential on the boundaries.

The Laplace equation can be transformed according to the method of separation of variables to address the multiple dimensional problem or the one-dimensional transient thermal analysis. Taking the case of transient thermal heat conduction, the temperature is assumed to be a function of space and time and can be expressed as

$$T(x, t) = T(x)\Gamma(t)$$

Substitution of the function into the heat equation, the form can be represented as

$$\frac{1}{T(x)}\frac{\mathrm{d}^2T(x)}{\mathrm{d}x^2}=\frac{1}{\alpha\Gamma}\frac{\mathrm{d}\Gamma}{\mathrm{d}t}=-\lambda^2$$

Hence, we can solve the spatial variation independently, by solving the solution for the second-order form of the modified Laplace equation as

$$\frac{\mathrm{d}^2T(x)}{\mathrm{d}x^2}+\lambda^2T(x)=0$$

Let us define a temperature and heat flux at a first boundary and second boundary. A solution of the equation by inspection, satisfies the form

$$T(x)=T^{\alpha}\frac{\sin(\lambda x)}{\sin(\lambda\Delta)}-T^{\beta}\frac{\sin\{\lambda(x-\Delta)\}}{\sin(\lambda\Delta)}$$

where the first surface β is at $x=0$ and the second surface α is at $x=\Delta$.

The derivative of temperature is

$$\frac{\mathrm{d}}{\mathrm{d}x}T(x)=T^{\alpha}\frac{\cos(\lambda x)}{\sin(\lambda\Delta)}-T^{\beta}\frac{\cos\{\lambda(x-\Delta)\}}{\sin(\lambda\Delta)}$$

From this relationship, we can form a forward matrix relation of the derivative of temperature at the boundaries to the temperature at the boundaries.

$$\begin{bmatrix}\left.\dfrac{\mathrm{d}T}{\mathrm{d}x}\right|^{\alpha}\\ \left.\dfrac{\mathrm{d}T}{\mathrm{d}x}\right|^{\beta}\end{bmatrix}=\lambda\begin{bmatrix}\cot(\lambda\Delta) & \dfrac{-1}{\sin(\lambda\Delta)}\\ \dfrac{1}{\sin(\lambda\Delta)} & -\cot(\lambda\Delta)\end{bmatrix}\begin{bmatrix}T^{\alpha}\\ T^{\beta}\end{bmatrix}$$

From the relationship of heat flux to temperature we can represent this in the form

$$\begin{bmatrix}q^{\alpha}\\ q^{\beta}\end{bmatrix}=-\lambda k\begin{bmatrix}\cot(\lambda\Delta) & \dfrac{-1}{\sin(\lambda\Delta)}\\ \dfrac{1}{\sin(\lambda\Delta)} & -\cot(\lambda\Delta)\end{bmatrix}\begin{bmatrix}T^{\alpha}\\ T^{\beta}\end{bmatrix}$$

This form allows for the representation of the heat flux from the temperature at the boundaries. Representing the temperature as a function of the heat flux

$$\begin{bmatrix} T^{\alpha} \\ T^{\beta} \end{bmatrix} = -\frac{1}{\lambda k} \begin{bmatrix} \cot(\lambda\Delta) & \frac{-1}{\sin(\lambda\Delta)} \\ \frac{1}{\sin(\lambda\Delta)} & -\cot(\lambda\Delta) \end{bmatrix} \begin{bmatrix} q^{\alpha} \\ q^{\beta} \end{bmatrix}$$

From this matrix approach, a region can be represented as n regions of a uniform property. The solution to the stratified medium can be solved using matrix multiplication. From an electrical circuit analogy, each physical film can be thought of as a two-port film where the information is zero at the terminals. Note that the electrical and magnetic potentials and fluxes can also utilize this methodology. This methodology can be used for the transient one-dimensional solution or oscillatory signals that are separable according to the form required for separation of variables. In these cases, the variable separation represents different physical meaning, but mathematically reduces to the form above. This methodology is extendable to cylindrical and spherical systems where the temperature and heat flux information is solved at the boundaries. The cylindrical or spherical annulus produces a matrix which is the solution of the Laplace equation in the physical region.

2.1.4 Heat Equation with Variable Conductivity

During an ESD event, the temperature transition varies from the ambient temperature to melting temperatures. In the Wunsch–Bell model, it is assumed that the thermal conductivity and specific heat–mass density product are constant. From the equation of linear heat flow, where the thermal conductivity is not a constant, let

$$\frac{\partial}{\partial x}\left(K\frac{\partial T}{\partial x}\right) - \rho c_{\mathrm{p}}\frac{\partial T}{\partial t} = 0$$

From this form, let us differentiate the terms

$$\frac{\partial}{\partial x}\left(K\frac{\partial T}{\partial x}\right) - \rho c_{\mathrm{p}}\frac{\partial T}{\partial t} = 0$$

where

$$\frac{\partial K}{\partial T} = K\left\{\frac{d(\ln K)}{dT}\frac{\partial T}{\partial x}\right\}$$

Then the heat equation with variable conductivity can be put in the form

$$K\frac{\partial^2 T}{\partial x^2} + K\frac{\mathrm{d}(\ln K)}{\mathrm{d}T}\left(\frac{\partial T}{\partial x}\right)^2 - \rho c_{\mathrm{p}}\frac{\partial T}{\partial t} = 0$$

2.1.4.1 Kirchoff transformation

Using the Kirchoff transformation, where

$$\Theta = \int_0^T \frac{K(T')}{K_0} \mathrm{d}T'$$

the heat equation with variable conductivity can be put into a simplified form. In this form, the transformed variable is equal to zero when $T = 0$.

From this expression, the partial derivatives with respect to space and time of the new variable are

$$\frac{\partial}{\partial t}\Theta = \frac{\partial}{\partial t}\int_0^T \frac{K(T')}{K_0} \mathrm{d}T'$$

From Leibniz's theorem, the variable of time can be exchanged with integration as

$$\frac{\partial}{\partial t}\Theta = \frac{\partial}{\partial t}\int_0^T \frac{K(T')}{K_0} \mathrm{d}T' = \int_0^T \frac{\partial}{\partial t}\left(\frac{K(T')}{K_0}\right)\mathrm{d}T' + \left(\frac{K(T)}{K_0}\right)\frac{\mathrm{d}T}{\mathrm{d}t} - \left(\frac{K_0}{K_0}\right)\frac{\mathrm{d}(T_0)}{\mathrm{d}t}$$

In this expression, the first and third terms are zero, hence

$$\frac{\partial}{\partial t}\Theta = \frac{\partial}{\partial t}\int_0^T \frac{K(T')}{K_0} \mathrm{d}T' = \left(\frac{K(T)}{K_0}\right)\frac{\partial T}{\partial t}$$

Applying Leibniz's theorem for the first and second partial derivatives as a function of space, we obtain

$$\frac{\partial}{\partial x}\Theta = \frac{\partial}{\partial x}\int_0^T \frac{K(T')}{K_0} \mathrm{d}T' = \int_0^T \frac{\partial}{\partial x}\left(\frac{K(T')}{K_0}\right)\mathrm{d}T' + \left(\frac{K(T)}{K_0}\right)\frac{\mathrm{d}T}{\mathrm{d}x} - \left(\frac{K(0)}{K_0}\right)\frac{\mathrm{d}(T_0)}{\mathrm{d}x}$$

In this expression, the third term is zero. Reapplying Leibniz's theorem, we obtain

$$\frac{\partial^2}{\partial x^2}\Theta = \frac{\partial}{\partial x}\left\{\frac{\partial}{\partial x}\int_0^T \frac{K(T')}{K_0} \mathrm{d}T'\right\} = \frac{\partial}{\partial x}\left\{\int_0^T \frac{\partial}{\partial x}\left(\frac{K(T')}{K_0}\right)\mathrm{d}T' + \left(\frac{K(T)}{K_0}\right)\frac{\mathrm{d}T}{\mathrm{d}x}\right\}$$

$$\frac{\partial^2}{\partial x^2}\Theta = \frac{\partial}{\partial x}\left\{\frac{K(T)}{K_0}\frac{\mathrm{d}T}{\mathrm{d}x}\right\} + \frac{\partial}{\partial x}\left\{\int_0^T \frac{\partial}{\partial x}\left(\frac{K(T')}{K_0}\right)\mathrm{d}T'\right\}$$

From the first term

$$\frac{\partial}{\partial x}\left(\frac{K(T)}{K_0}\frac{\partial T}{\partial x}\right) = \frac{K(T)}{K_0}\frac{\partial^2 T}{\partial x^2} + \frac{1}{K_0}\frac{\partial K}{\partial T}\left(\frac{\mathrm{d}T}{\mathrm{d}x}\right) = \frac{K(T)}{K_0}\frac{\partial^2 T}{\partial x^2} + \frac{K(T)}{K_0}\left\{\frac{\mathrm{d}(\ln K)}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}x}\right\}\left(\frac{\mathrm{d}T}{\mathrm{d}x}\right)$$

Substituting the transformed variables into the heat equation, we obtain the heat equation with variable coefficients in the form of

$$\frac{\partial^2 \Theta}{\partial x^2} - \frac{1}{\kappa}\frac{\partial \Theta}{\partial t} = 0$$

where we express the transformed temperature thermal diffusivity as

$$\kappa = \frac{K}{\rho c}$$

For ESD events, the transformed temperature formulation of the heat equation proposed by Kirchoff is valuable for power to failure prediction and model development. Ash showed that utilizing a form of the Kirchoff formulation, the time dependence of the Wunsch–Bell relations is maintained in the thermal diffusion regime for certain forms of thermal conductivity temperature dependence. Using the correct temperature dependence will allow for improved prediction of the effective area factor for ESD power to failure models.

2.1.4.2 Boltzmann transformation

The heat conduction equation is a second-order partial differential equation which is second order in space and first order in time. Assuming a new variable which is a function of both space and time, it can be shown that the second-order partial differential equation can be modified as a second-order ordinary differential equation. With this conversion, solutions can be simplified using ordinary differential equation methods.

Let a function be

$$\xi = \frac{x}{\sqrt{t}}$$

Then the partial differential of space and time can be converted to

$$\frac{\partial}{\partial x} = \frac{1}{\sqrt{t}}\frac{\partial}{\partial \xi} \quad \text{and} \quad \frac{\partial}{\partial t} = -\frac{1}{2}\xi\frac{1}{t}\frac{\partial}{\partial \xi}$$

From this form, the heat conduction equation can substitute in the above operators

$$\frac{\partial}{\partial x}\left(K\frac{\partial T}{\partial x}\right) - \rho c\frac{\partial T}{\partial t} = 0$$

where

$$\left(\frac{1}{\sqrt{t}}\frac{\partial}{\partial \xi}\right)\left\{K\left(\frac{1}{\sqrt{t}}\frac{\partial T}{\partial \xi}\right)\right\} - \rho c\left\{-\left(\frac{1}{2}\xi\frac{1}{t}\right)\frac{\partial T}{\partial \xi}\right\} = 0$$

In this form, the equation reduces to an ordinary differential equation of second order with variable coefficients.

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left(K\frac{\mathrm{d}T}{\mathrm{d}\xi}\right)+\rho c\xi\frac{\mathrm{d}T}{\mathrm{d}\xi}=0$$

Applying the Boltzmann transformation to the Kirchoff formulation,

$$\frac{\partial^2\Theta}{\partial x^2}-\frac{1}{\kappa}\frac{\partial\Theta}{\partial t}=0$$

Substituting in the Boltzmann transformation operators

$$\left\{\left(\frac{1}{\sqrt{t}}\frac{\partial}{\partial\xi}\right)\left(\frac{1}{\sqrt{t}}\frac{\partial}{\partial\xi}\right)-\frac{1}{\kappa}\left(-\frac{1}{2}\xi\frac{1}{t}\frac{\partial}{\partial\xi}\right)\right\}\Theta=0$$

or

$$\kappa\frac{\mathrm{d}^2\Theta}{\mathrm{d}^2\xi}+\frac{1}{2}\xi\frac{\mathrm{d}\Theta}{\mathrm{d}\xi}=0$$

In this form, the heat equation utilizes the Kirchoff formulation of variables and the Boltzmann transformation, the equation is reduced to an ordinary differential equation with variable coefficients and allows for a variable thermal conductivity. From the form of this equation, we can verify that the solution to the equation is the error function.

Let us assume a solution of the form

$$\Theta(\xi)=A\ erf\left\{\frac{\xi}{2\sqrt{\kappa}}\right\}$$

where

$$erf\left\{\frac{\xi}{2\sqrt{\kappa}}\right\}=\frac{2}{\sqrt{\pi}}\int_0^{\frac{\xi}{2\sqrt{\kappa}}}\exp\{-t^2\}\mathrm{d}t$$

Applying Leibniz's theorem of differentiation, we obtain

$$\begin{aligned}\frac{\mathrm{d}\Theta}{\mathrm{d}\xi}&=\frac{2}{\sqrt{\pi}}\left\{\int_0^{\frac{\xi}{2\sqrt{\kappa}}}\frac{\mathrm{d}}{\mathrm{d}\xi}\{\exp(-t^2)\}\mathrm{d}t+\exp\left\{-\left(\frac{\xi}{2\sqrt{\kappa}}\right)^2\right\}\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{\xi}{2\sqrt{\kappa}}\right)\right\}\\&=\frac{2}{\sqrt{\pi}}\frac{1}{2\sqrt{\kappa}}\exp\left\{-\frac{\xi^2}{4\kappa}\right\}\end{aligned}$$

then

$$\frac{\mathrm{d}^2\Theta}{\mathrm{d}\xi^2}=\frac{2}{\sqrt{\pi}}\frac{1}{2\sqrt{\kappa}}\frac{-2\xi}{4\kappa}\exp\left\{-\frac{\xi^2}{4\kappa}\right\}$$

Substitution of the derivatives into the expression shows that the error function in this form is a general solution of the Kirchoff representation of the heat conduction equation with conductivity which is a function of temperature.

2.1.5 Duhamel Formulation

For the evaluation of the power to failure of a semiconductor, a criterion using a form of the Duhamel formulation can be applied [4]. The importance of this relationship is valuable for defining the failure time, as well as waveform conversion from one pulse waveform to another.

The Duhamel principle for the heat equation is given a heat equation relationship in a domain D, with a boundary C, where the temperature on the boundary is zero and an initial temperature field of $u(x,y,z,\ t=0)=f(x,y,z)$, and a second solution $w(x,y,z,t)$ with $w=0$ on boundary C, and $w(x,y,z,t=0)=0$ with an inhomogeneous heat equation with drive term $h(x,y,z,t)$, then a linear operator exists where $u(x,y,z,t)=L\,[f(\xi,\ \eta,\ \zeta)]\,(x,y,z,t)$ where

$$w(x,y,z,t)=\int_0^t L[h(\xi,\eta,\zeta,\tau)](x,y,z,t-\tau)\mathrm{d}\tau$$

Tasca first applied this to explain the power to failure of semiconductors. Assuming an arbitrary input power waveform $P(t)$ and a damage threshold power $P_{\mathrm{D}}\ (t_{\mathrm{p}})$ associated with a rectangular pulse of pulse width t_{p}, the criterion for damage is when the following inequality is satisfied

$$\int_0^t P(\tau)\left\{\frac{\partial}{\partial(t-\tau)}\left[\frac{1}{P_D(t-\tau)}\right]\right\}\mathrm{d}\tau\geq 1$$

Then the time at which it is satisfied is the time when the critical time is achieved. Hence we can define the critical time according to the relationship, where t_{c} is the critical time, or the time to failure $t=t_{\mathrm{f}}$.

$$\int_0^{tf} P(\tau)\left\{\frac{\partial}{\partial(t-\tau)}\left[\frac{1}{P_D(t-\tau)}\right]\right\}\mathrm{d}\tau=1$$

From the relationship

$$T(\underline{r},t)=T_0+\int_0^t P(\tau)\frac{\mathrm{d}}{\mathrm{d}(t-\tau)}\{H(r,t-\tau)\}\mathrm{d}\tau$$

Given that the power is independent of time, then

$$T(\underline{r},t)=T_0+P_0H(r,t)$$

From this expression, we can define the power to failure, and the failure time as

$$P_0(t_\mathrm{f}) = \frac{T_\mathrm{c} - T_0}{H(t_\mathrm{f})}$$

Hence

$$H(t_\mathrm{f}) = \frac{T_\mathrm{c} - T_0}{P_0(t_\mathrm{f})}$$

From this relationship, we can relate power and temperature as an implicit form from the critical failure time, and eliminate the geometric information contained in $H(r,t)$

$$T(\underline{r}, t) = T_0 + \int_0^t P(\tau) \frac{\mathrm{d}}{\mathrm{d}(t-\tau)} \left\{ \frac{T_\mathrm{c} - T_0}{P_0(t-\tau)} \right\} \mathrm{d}\tau$$

This relationship is important for evaluation of an ESD event for transmission line pulse systems, in that we can relate the information of power, time, and critical temperature of the medium without implicit understanding of the spatial information. Normalizing the expression, we obtain a normalized expression which contains no spatial information

$$\frac{T(t) - T_0}{T_\mathrm{c} - T_0} = \int_0^t P(\tau) \frac{\mathrm{d}}{\mathrm{d}(t-\tau)} \left\{ \frac{1}{P_0(t-\tau)} \right\} \mathrm{d}\tau$$

Hence we can define a criterion that when the temperature is at the critical temperature then, the time to failure can be defined as

$$\int_0^{tf} P(\tau) \frac{\mathrm{d}}{\mathrm{d}(t-\tau)} \left\{ \frac{1}{P_0(t-\tau)} \right\} \mathrm{d}\tau = 1$$

In many applications, it is assumed that the critical temperature is associated with the intrinsic temperature T_i. Hence we can define a criteria where the $T_\mathrm{c} = T_\mathrm{i}$ or

$$\frac{T(t) - T_0}{T_i - T_0} = \int_0^t P(\tau) \frac{\mathrm{d}}{\mathrm{d}(t-\tau)} \left\{ \frac{1}{P_0(t-\tau)} \right\} \mathrm{d}\tau$$

Applying the failure criteria to semiconductors, for very short pulses, in the adiabatic regime, we can express the power to failure for a rectangular pulse of where

$$P_\mathrm{D}(t_\mathrm{p}) = \frac{MC_\mathrm{p}(T_f - T_0)}{t_\mathrm{p}}$$

then

$$\frac{\partial}{\partial(t-\tau)} \left[\frac{1}{P_\mathrm{D}(t-\tau)} \right] = \frac{\partial}{\partial(t-\tau)} \left[\frac{t-\tau}{MC_\mathrm{p}(T_\mathrm{f} - T_0)} \right] = \frac{1}{MC_\mathrm{p}(T_\mathrm{f} - T_0)}$$

Then for an arbitrary power profile $P(t)$ where the pulse is short relative to the thermal diffusion time in the medium, we have a general failure criteria in the adiabatic regime for the failure time as

$$\int_0^t \frac{P(\tau)}{MC_{\mathrm{p}}(T_{\mathrm{f}} - T_0)} \mathrm{d}\tau \geqslant 1$$

In the thermal diffusion regime, let us assume a form of

$$P_{\mathrm{D}}(t_{\mathrm{p}}) = \frac{C(T_{\mathrm{f}} - T_0)}{\sqrt{t_{\mathrm{p}}}} \qquad C = A\sqrt{\pi K \rho C_{\mathrm{p}}}$$

then

$$\frac{\partial}{\partial(t-\tau)}\left[\frac{1}{P_{\mathrm{D}}(t-\tau)}\right] = \frac{\partial}{\partial(t-\tau)}\left[\frac{(t-\tau)^{1/2}}{C(T_{\mathrm{f}} - T_0)}\right] = \frac{1}{2C(T_{\mathrm{f}} - T_0)}\frac{1}{\sqrt{t-\tau}}$$

Substituting into the integral expression we obtain a general expression for the time to failure for an arbitrary power profile $P(t)$

$$\int_0^t \frac{P(\tau)}{2C(T_{\mathrm{f}} - T_0)\sqrt{t-\tau}} \mathrm{d}\tau = \int_0^t \frac{1}{2\sqrt{\pi K \rho C_{\mathrm{p}}}(T_{\mathrm{f}} - T_0)} \frac{P(\tau)}{\sqrt{t-\tau}} \mathrm{d}\tau \geqslant 1$$

2.2 ELECTROTHERMAL MODELS

In the next sections, the methods will be applied to well-established ESD models that were established in the 1960s to 1990s. The base of these ESD models are solutions of the thermal diffusion equation, and the establishment of a time constant hierarchy relative to the thermal response of the system and the width of the applied power pulse.

2.2.1 Tasca Model

Tasca addressed the problem of thermal breakdown by assuming that thermal breakdown initiates from a single current constriction. In the model proposed by Tasca [4], it was assumed that all the junction current passed through the spherical defect. As a second assumption in the model, it was assumed that the severity of the defective site was such that only heat is dissipated to the surrounding medium. Tasca assumed that the defect where breakdown occurred was spherical in nature. The thermal conductivity and specific heat were assumed a constant. The model assumed a finite sphere of radius a. The sphere was assumed to exist in an infinite medium.

Solving for the problem of a spherical region in an infinite medium, the temperature T_c of the current constriction is

$$T_c - T_a = \frac{a^2 Q}{2K} F^*$$

where Q is the heat rate, a is the radius of the spherical current constriction, K is the thermal conductivity, and F^* is a function where

$$F^* = F_1 \qquad R = 0$$

$$F^* = F_2 \qquad 0 < R < a$$

The solution for F_1 can be represented as

$$F_1 = 1 + \left(\frac{2kt}{a^2} - 1\right) erf\left(\frac{a}{2\sqrt{kt}}\right) - 2\sqrt{\frac{kt}{\pi a^2}} \exp\left(-\frac{a^2}{4\sqrt{kt}}\right)$$

and

$$F_2 = \frac{2kt}{a^2}\left\{1 - \frac{2a}{R} i^2 erfc\left(\frac{a-R}{2\sqrt{kt}}\right) + \frac{2a}{R} i^2 erfc\left(\frac{a+R}{2\sqrt{kt}}\right) - \frac{4\sqrt{kt}}{R} i^3 erfc\left(\frac{a-R}{2\sqrt{kt}}\right) + \frac{4\sqrt{kt}}{R} i^3 erfc\left(\frac{a+R}{2\sqrt{kt}}\right)\right\}$$

The solution represents an image charge of equal and opposite magnitude about the spherical surface at $R = a$.

Tasca showed that this expression can be reduced to a simpler form where the function F^* is solved at a point near the radius a, where F^* is approximated as

$$F^* = \frac{\dfrac{4Kt}{a^2 \rho c}}{2 + 3\sqrt{\dfrac{4Kt}{a^2 \rho c}} + \dfrac{4Kt}{a^2 \rho c}}$$

From this expression, we can substitute in order to solve for temperature

$$T_c - T_a = \frac{a^2 Q}{2K}\left\{\frac{\dfrac{4Kt}{a^2 \rho c}}{2 + 3\sqrt{\dfrac{4Kt}{a^2 \rho c}} + \dfrac{4Kt}{a^2 \rho c}}\right\}$$

Let us define the heat rate per unit volume per unit time as

$$Q = \frac{P}{\frac{4}{3}\pi a^3}$$

Substituting for this expression we can obtain a relationship between the temperature and power

$$T_c - T_a = \frac{a^2}{2K}\left(\frac{P}{\frac{4}{3}\pi a^3}\right)\left\{\frac{\frac{4Kt}{a^2\rho c}}{2+3\sqrt{\frac{4Kt}{a^2\rho c}}+\frac{4Kt}{a^2\rho c}}\right\}$$

Note, from the expression of the relationship that temperature is equal to the product of the thermal impedance and power, we obtain

$$R_{TH} = \frac{a^2}{2K}\left(\frac{1}{\frac{4}{3}\pi a^3}\right)\left\{\frac{\frac{4Kt}{a^2\rho c}}{2+3\sqrt{\frac{4Kt}{a^2\rho c}}+\frac{4Kt}{a^2\rho c}}\right\}$$

Solving for power to failure, and substituting in a spherical volume Δ, radius r and surface area S, from

$$T_c - T_a = \frac{a^2}{2K}\left(\frac{P}{\frac{4}{3}\pi a^3}\right)\left\{\frac{\frac{4Kt}{a^2\rho c}}{2+3\sqrt{\frac{4Kt}{a^2\rho c}}+\frac{4Kt}{a^2\rho c}}\right\}$$

$$P = \left(\frac{2K}{a^2}\right)\left(\frac{a^2\rho C}{4Kt}\right)\left(\frac{4}{3}\pi a^3\right)\left\{2+3\sqrt{\frac{4Kt}{a^2\rho c}}+\frac{4Kt}{a^2\rho c}\right\}(T_c - T_a)$$

$$P = \left\{\frac{\rho C\Delta}{t}+3\left(\frac{\rho C}{2t}\right)\left(\frac{4\pi a2}{3}\right)a\sqrt{\frac{4Kt}{a^2\rho c}}+\left(\frac{2\rho C}{3t}\pi a^3\right)\frac{4Kt}{a^2\rho c}\right\}(T_c - T_a)$$

Hence

$$P = \left\{\frac{\rho C\Delta}{t}+S\left(\frac{\rho CK}{t}\right)^{1/2}+\frac{8}{3}\pi aK\right\}(T_c - T_a)$$

Then power is the sum of the transient power stored in the spherical volume Δ, the transient power that diffuses through its surface S, and the steady state power. The power to failure is when the time is the critical time to failure, where

$$P_{\mathrm{f}} = \left\{ \frac{\rho C_{\mathrm{p}} \Delta}{t_{\mathrm{f}}} + S\left(\frac{K \rho C_{\mathrm{p}}}{t_{\mathrm{f}}} \right)^{1/2} + \frac{8\pi K r}{3} \right\} (T_{\mathrm{c}} - T_0)$$

At pulse widths significantly less than the thermal diffusion time, the second and third terms are negligible compared with the first term. The first term represents the adiabatic term. For pulse widths of the order of the thermal diffusion time, the second term dominates. As the pulse widths become significantly greater than the thermal diffusion time, the expression is dominated by the third term, which is time independent.

The power to failure model of the spherical defect is very dependent on the geometry of the defect as well as the variables for the volume, surface area and radius. Assuming we can verify that in fact the defect is spherical in nature, and can determine the critical temperature of failure, it is then possible to extract the size of the defect. For very short times, the spherical volume can be predicted as

$$\Delta = \left\{ \frac{P_{\mathrm{f}} t_{\mathrm{f}}}{\rho C_{\mathrm{p}} (T_{\mathrm{c}} - T_0)} \right\}$$

where a plot of volume can be calculated based on the critical temperature to failure, the known heat capacity term, and the input power and pulse width. To calculate the surface area, a second term can be calculated as well, knowing the material properties

$$S = \left\{ \frac{P_{\mathrm{f}} (t_{\mathrm{f}})^{1/2}}{(K \rho C_{\mathrm{p}})^{1/2} (T_{\mathrm{c}} - T_0)} \right\}$$

By varying the pulse width of the applied pulse, the geometrical characteristics of the defect can be determined.

2.2.2 Wunsch–Bell Model

Wunsch and Bell analyzed the power to failure of single-component semiconductors [5]. In this work, they noted a few major effects. The study examined the power to failure of commercial single components as a function of pulse width. First, the power to failure of a forward-biased component was higher than that of reverse-biased components. Second, the power to failure decreased as the length of the applied pulse width increased. Third, the power to failure pulse width dependence was the same for positive and negative slopes. From this work, it was postulated that the failure mechanism was the same, independent of the bias polarity, but the magnitude of the power to failure is different. In the analysis of the Wunsch–Bell model (Figure 2.4), the time constant hierarchy compares the relationship with the applied pulse

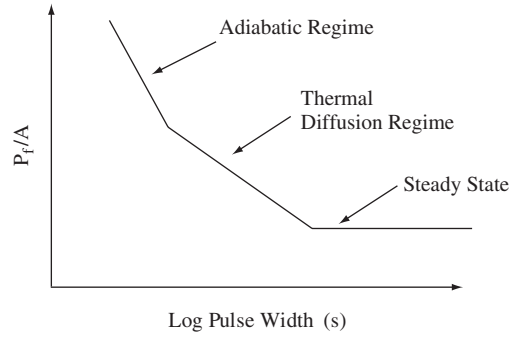

Figure 2.4 Wunsch–Bell model

From the thermal diffusion equation (heat equation) in the time-dependent form, the one-dimensional heat equation can be expressed as

$$\frac{\partial}{\partial x}\left(\kappa\frac{\partial T}{\partial x}\right) = \rho c_{\mathrm{p}}\frac{\partial T}{\partial t}$$

Assuming a uniform thermal conductivity, the equation can be put in the Laplacian form and expressed as

$$\frac{\partial^2 T}{\partial x^2} - \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0$$

where α is the thermal diffusivity. To maintain a one-dimensional analysis, we can assume a plane of thermal heat, with a thermal impulse Q where Q is a thermal pulse of finite pulse width. Solving the equation for temperature

$$T(x,t')|x,x',t = \frac{Q}{(\sqrt{4\pi Qt})}\exp\left\{-(x-x')^2/4\alpha t'\right\}$$

By integration over time

$$T(x,x')|x,x' = \frac{Q}{\left(\sqrt{4\pi\alpha}\right)}\int_0^t \frac{\exp\left\{-(x-x')^2/4\alpha(t-t')\right\}}{\sqrt{t-t'}}\mathrm{d}t'$$

Then

$$T(x-x') = T(x,x') = q\sqrt{\frac{t}{\pi\alpha}}\exp\{-(x-x')/4\alpha t\} - \frac{q|x-x'|}{2\alpha}erfc\left\{\frac{|x-x'|}{2\sqrt{\alpha t}}\right\}$$

Letting the variable $x - x' = 0$

$$T(0) = q\sqrt{\frac{t}{\pi\alpha}}$$

and letting

$$Q = \frac{P}{A}\frac{1}{\rho c_p}$$

where P is power, A is area, and substituting in for the thermal diffusivity, $1/\alpha = \rho c_p/\kappa$ one obtains the relationship

$$T - T_0 = \left(\frac{P}{A}\frac{1}{\rho c_p}\right)\sqrt{\frac{t}{\pi\alpha}} = \left(\frac{P}{A}\frac{1}{\rho c_p}\right)\sqrt{\frac{t}{\pi\frac{\kappa}{\rho c_p}}} = \left(\frac{P}{A}\right)\sqrt{\frac{t}{\pi(\kappa\rho c_p)}}$$

From this form, the increase in temperature can be calculated as a function of the input power, the pulse width, area and thermal properties. From this development, it can be seen that temperature increases linearly with the power. Note that this expression also shows that the power is the absorbed power. This analysis assumes that there is no reflected or transmitted power.

Let us assume that the structure fails at the intrinsic temperature T_i. Then the intrinsic temperature, can be expressed as

$$T_i - T_0 = \left(\frac{P_i}{A}\right)\sqrt{\frac{t}{\pi(\kappa\rho c_p)}}$$

where we define a new definition of the intrinsic power. The definition for the intrinsic power is the power needed to increase a medium to the intrinsic temperature for a given pulse width t.

Solving for the intrinsic power,

$$\frac{P_i}{A} = \frac{\sqrt{\pi\kappa\rho c_p}}{\sqrt{t}}(T_i - T_0)$$

When the temperature is the failure temperature, or maximum temperature, the relationship is

$$T_{max} - To = \left(\frac{P_f}{A}\right)\sqrt{\frac{t}{\pi(\kappa\rho c_p)}}$$

where the power at the maximum temperature is the power to failure, or

$$\frac{P_f}{A} = \frac{\sqrt{\pi\kappa\rho c_p}}{\sqrt{t}}(T_{max} - T_0)$$

The power to failure is associated with this maximum temperature. In many cases, the melting temperature is the maximum temperature associated with failure. In a complex structure or system, this may not always be the case.

In this analysis, the assumption is that the heat source is a one-dimensional plane wave thermal source propagating into the system with no thermal reflection loss and no thermal transmission out of the system. The development also assumes no internal generation mechanisms within the volume or space of interest. The equation also inherently assumes that the order of the pulse width of the thermal source is of the order of the thermal diffusion time. Given that the pulse width of the thermal source was significantly faster than the thermal diffusion process, the Laplacian term would be set to zero.

In short time scales, the one-dimensional time-dependent heat equation is

$$A\frac{\partial}{\partial x}\left(\kappa\frac{\partial T}{\partial x}\right)+Q(t)=A\rho c_{\mathrm{p}}\frac{\partial T}{\partial t}$$

Assuming that the thermal conductivity is a constant in the direction of heat flow

$$A\kappa\frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right)+Q(t)=A\rho c_{\mathrm{p}}\frac{\partial T}{\partial t}$$

and letting

$$Q(t')=\left\{\begin{array}{ll} q & 0<t'<t \\ 0 & t'>t \end{array}\right\}$$

In the hierarchy of time constants, when the pulse width is significantly shorter than thermal diffusion time, the equation reduces to a spatially independent heat equation where

$$\frac{\partial T}{\partial t}=\frac{Q(t)}{A\rho c_{\mathrm{p}}}$$

In this form, the equation is a first-order linear differential equation. Integrating the equation from an initial time $t'=0$ to the pulse length $t'=t$

$$\int_{t'=0}^{t'=t}\mathrm{d}t'\frac{\partial T}{\partial t'}=\int_{t'=0}^{t'=t}\mathrm{d}t'\frac{Q(t')}{A\rho c_{\mathrm{p}}}$$

From the total differential of temperature

$$\mathrm{d}T=\frac{\partial T}{\partial t}\mathrm{d}t$$

substituting for the integral expression where $t'=0$ the system is at ambient temperature $T'=T_{\mathrm{amb}}$, and at time $t'=t$, the temperature is $T'=T$

$$\int_{T'=T_{\mathrm{amb}}}^{T'=T}\mathrm{d}T=\int_{t'=0}^{t'=t}\mathrm{d}t'\frac{\partial T}{\partial t'}=\int_{t'=0}^{t'=t}\mathrm{d}t'\frac{Q(t')}{A\rho c_{\mathrm{p}}}$$

where

$$T - T_{\text{amb}} = \int_{t'=0}^{t'=t} \frac{Q(t')}{A\rho c_{\text{p}}} \mathrm{d}t' = \frac{q}{\rho c_{\text{p}} A} t$$

Let $q = P/A$, then

$$T - T_{\text{amb}} = \frac{P}{A} \frac{1}{A\rho c_{\text{p}}} t$$

In this time regime, the temperature is linearly proportional to the power absorbed in the medium, and is a function of the pulse width. The product of the power per unit area and the time is the energy input into the medium per unit area. When the temperature is the failure temperature $T = T_{\text{f}}$, then the power is power to failure $P = P_{\text{f}}$.

$$\frac{P_{\text{f}}}{A} = \frac{\rho c_{\text{p}} (T_{\text{f}} - T_{\text{amb}})}{t} A$$

2.2.3 Smith–Littau Model

Smith and Littau developed a model for the prediction of resistor elements [6]. In the Smith and Littau model, a resistor was modeled as a three-region structure. The thin-film resistor is present on an insulating film of dielectric representing an interlevel dielectric film where the thin-film resistor and interlevel dielectric interface is at $x = 0$. The interlevel dielectric film of film thickness x rests on a substrate wafer of thickness d, with the interlevel dielectric–substrate interface at x. In the model, it assumes that all heat loss is from thermal conduction (e.g. heat loss from radiation is negligible and net heat flux at the interface is zero).

From the heat diffusion equation,

$$\frac{\partial^2 T}{\partial x^2} - \frac{1}{\alpha} \frac{\partial T}{\partial t} = 0$$

The boundary condition at $x = 0$ is the boundary condition of the heat flux from the interconnect into the insulating interlevel dielectric.

$$q = -\kappa \frac{\partial T}{\partial x} \qquad x = 0$$

Let heat flux per unit area equal F, then we can express the first derivative at the interface as

$$\frac{\partial T}{\partial x} = -\frac{F}{\kappa} \qquad x = 0 \qquad \tau \geqslant 0$$

and the additional boundary conditions that at an infinite distance the temperature is zero from the initial time to infinite time, and that the substrate is a perfect heat sink with temperature zero at the interlevel dielectric–substrate interface.

$$x = \infty \qquad T = 0 \qquad \tau \geqslant 0$$

$$x = x \qquad T(x) = 0 \quad \tau \leqslant 0 \qquad 0 \leqslant x \leqslant \infty$$

The solution to the partial differential equation as a function of time and temperature can be expressed as

$$T(x,t) = 2C\sqrt{\alpha\tau u}\left(\frac{2}{\sqrt{\pi}}\int_0^u \exp\left(-u^2\right)\mathrm{d}u + \frac{\exp\left(-u^2\right)}{\sqrt{\pi u}} - 1\right)$$

where

$$u = \left(\frac{x}{2}\right)\sqrt{2\tau}$$

To satisfy the conduction boundary condition at the thin-film resistor to interlevel dielectric interface, let $x = 0$ where in the limit that x approaches zero, the variable u also approaches zero. Then

$$T(0,t) = 2C\sqrt{\alpha\tau u}\left(\frac{2}{\sqrt{\pi}}\lim_{x\to 0}\int_0^x \exp\left(-u^2\right)\mathrm{d}u + \lim_{u\to 0}\frac{\exp\left(-u^2\right)}{\sqrt{\pi u}} - 1\right)$$

$$\lim_{u\to 0} 2C\sqrt{\alpha\tau u}\frac{\exp\left(-u^2\right)}{\sqrt{\pi u}} = \lim_{u\to 0} 2C\sqrt{\alpha\tau}\frac{\exp\left(-u^2\right)}{\sqrt{\pi}} \to 2C\sqrt{\frac{\alpha\tau}{\pi}}$$

This can then be expressed as

$$T(0,t) = 2C\sqrt{\frac{\alpha t}{\pi}}$$

where $C = F/\kappa$, hence

$$T(0,t) = 2\frac{F}{\kappa}\sqrt{\frac{\alpha t}{\pi}}$$

The peak temperature in the system will occur at the interface of the thin-film resistor and the interlevel dielectric.

In the thin-film resistor, the current I can flow through the resistor element parallel to the interlevel dielectric film. The resistor element has length l, width w. The temperature is increased as a result of Joule heating, where the power per unit area of the resistor element is

$$P = I^2R = I^2\rho\frac{L}{Wt} = I^2\rho_{\mathrm{sq}}\frac{L}{W}$$

Calculating the power per unit area

$$\frac{P}{A} = \frac{1}{LW} I^2 \rho_{sq} \frac{L}{W} = \frac{I^2 \rho_{sq}}{W} = \frac{V^2}{\rho_{sq} L^2}$$

From this relationship, we can solve for the relationship between the temperature at the interface, the current, voltage and time

$$T(0,t) = 2\frac{I^2 \rho_{sq}}{KW^2}\sqrt{\frac{\alpha t}{\pi}}$$

$$T(0,t) = 2\frac{V^2}{K\rho_{sq} L^2}\sqrt{\frac{\alpha t}{\pi}}$$

The temperature at the interface is the peak temperature. Hence, when the temperature is the failure temperature, the current and voltage are the current to failure, and the voltage is the voltage to failure, for the pulse width $t = \tau$

$$T_f(0, t = \tau) = 2\frac{{I_f}^2 \rho_{sq}}{KW^2}\sqrt{\frac{\alpha \tau}{\pi}}$$

and

$$T_f(0, t = \tau) = 2\frac{{V_f}^2}{K\rho_{sq} L^2}\sqrt{\frac{\alpha \tau}{\pi}}$$

Substituting for the thermal diffusivity

$$T_f(0, t = \tau) = 2\frac{V_f^2}{K\rho_{sq} L^2}\sqrt{\frac{K\tau}{\rho C_p \pi}}$$

Using the thermal conductivity for the semi-insulating material in cal cm^{-1} s^{-1} and bringing K into the square root

$$T_f(0, t = \tau) = 2\frac{{V_f}^2}{4.184\rho_{sq} L^2}\sqrt{\frac{\tau}{\rho K C_p \pi}}$$

Solving for the failure voltage

$${V_f}^2 = \frac{4.184\rho_{sq} L^2 T_c}{2}\frac{\sqrt{\rho K C_p \pi}}{\sqrt{\tau}}$$

The Smith–Littau model assumes that the top of the resistor does not have any insulating material. A second solution assuming semi-insulating materials on both top and bottom is

$${V_f}^2 = 4.184\rho_{sq} L^2 T_c \frac{\sqrt{\rho K C_p \pi}}{\sqrt{\tau}}$$

This assumption assumes that the time scale is such that the boundary conditions on top and bottom appear identical.

2.2.4 Arkihpov–Astvatsaturyan–Godovosyn–Rudenko Model

In the Arkhipov–Astvatsaturyan–Godovsyn–Rudenko (AAGR) model [8], the defect is not regarded as a spherical region, but as a cylindrical region. Assuming that the cylindrical defect has radius b and diameter D, an expression for the long pulses prior to the steady state condition can be obtained as

$$P_{\mathrm{f}} = \frac{[4\pi Ka(T_{\mathrm{c}} - T_0)]}{\left[\ln\left(\frac{t_{\mathrm{f}}}{b^2/4\pi D}\right) + \ln\left(\frac{4}{\pi}\right)\right]}$$

The AAGR model assumes the cylindrical nature of the defect. Experimental evidence in GaAs and other compound semiconductors clearly shows single and multiple filamentation in the defective region where the cylindrical column is small in radius compare with the length of the column.

2.2.5 Vlasov–Sinkevitch Model

Vlasov and Sinkevitch [9] also developed a physical model to explain the defect region. They obtained an expression for the maximum temperature T at time t in a heat center after pulsed power P. When the defect reaches the critical temperature, the power to failure can be expressed as

$$P_{\mathrm{f}} = \frac{[\pi\lambda R(T_{\mathrm{c}} - T_0)]}{\left[1 - erf\left\{\frac{R}{2a\sqrt{t}}\right\} - \frac{2a\sqrt{t}}{R\sqrt{\pi}}\left\{\exp\left(-\frac{R^2}{4a^2t}\right) - 1\right\}\right]}$$

where

$$a^2 = \lambda/c\rho$$

where λ is the thermal conductivity. In this expression, the thermal diffusion time is $R/2a$. The ratio of this time $R/2a$ relative to the pulse time t that determines the nature of the denominator.

2.2.6 Dwyer–Franklin–Campbell Model

Dwyer, Franklin and Campbell [10] developed a physical thermal model of a planar device accounting for the three-dimensional nature of the structure. In many semiconductor devices, the scale length of the three physical dimensions are significantly different due to the significant difference in the scaling of structures in the horizontal

and vertical directions. Dwyer, Franklin and Campbell developed a model based on first principles, using a Green's function methodology where the solution was not a function of a specific power profile. The heat transfer equation is expressed as

$$\frac{\partial T}{\partial t} - D\nabla^2 T = \frac{q(t)}{\rho C_{\mathrm{p}}}$$

where $D = k/\rho C_{\mathrm{p}}$. In the DFC model, it was assumed that the rate of heating inside the volume is constant and zero outside. For a constant output power, P_0, in volume Δ this can be put into a Poisson equation form,

$$\nabla^2 T = -\frac{P_0}{D\rho C_p \Delta}$$

The general three dimensional solution of the Poisson equation can be expressed as

$$T(\underline{r}) = T_0 + \frac{P_0}{K\Delta}\int \frac{d\underline{r}'}{4\pi|\underline{r} - \underline{r}'|}$$

The Green's function solution to the heat equation is the response from a Dirac delta function which is a function of space and time. The expression is

$$\frac{\partial G}{\partial t} - D\nabla^2 G = \delta(\underline{r} - \underline{r}')\delta(t - \tau)$$

The Green's function solution is the three-dimensional Gaussian expression

$$G(r, r', t, \tau) = G(r - r', t - \tau) = \frac{1}{[4\pi D(t-\tau)]^{3/2}} \exp\left\{-\frac{[\underline{r} - \underline{r}']^2}{4D(t-\tau)}\right\}$$

Solving for temperature at position r, and time t,

$$T(\underline{r}, t) = T_0 + \iint P(\tau) G(r - r', t - \tau) \mathrm{d}r' \mathrm{d}\tau$$

The integrals can be separated and expressed as

$$T(\underline{r}, t) = T_0 + \int_0^t \frac{P(\tau)}{\rho Cp\Delta} \mathrm{d}\tau \int_\Delta \frac{\mathrm{d}\underline{r}'}{[4\pi D(t-\tau)]^{3/2}} \exp\left\{-\frac{(\underline{r} - \underline{r}')^2}{4D(t-\tau)}\right\}$$

This expression can be expressed in terms of the complementary error function *erfc*(*r*,*t*). Let

$$H(\underline{r}, t) = \frac{1}{4\pi K\Delta} \int \frac{erfc\left\{\frac{|\underline{r} - \underline{r}'|}{2\sqrt{Dt}}\right\}}{|\underline{r} - \underline{r}'|} \mathrm{d}\underline{r}'$$

then we can express the temperature–power relationship as

$$T(\underline{r},t) = T_0 + \int_0^t P(\tau)\frac{\mathrm{d}}{\mathrm{d}(t-\tau)}H(\underline{r},t-\tau)\mathrm{d}\tau$$

The above relationship is the Duhamel formula.

From this expression, if the pulse power is independent of time, then the power term can be removed from the integral expression, and

$$T(\underline{r},t) = T_0 + P_0\int_{\tau=0}^{\tau=t}\frac{\mathrm{d}}{\mathrm{d}(t-\tau)}H(\underline{r},t-\tau)\mathrm{d}\tau$$

Letting $\chi = t-\tau$, the integral can be transformed from the lower limit of $\chi = t$ to upper limit of $\chi = 0$, and $\mathrm{d}\chi = \mathrm{d}\,(t-\tau) = -\mathrm{d}\,\tau$

$$T(\underline{r},t) = T_0 + P_0\int_{x=t}^{x=0}\frac{\mathrm{d}}{\mathrm{d}(x)}H(\underline{r},x)(-\mathrm{d}x)$$

then

$$T(\underline{r},t) = T_0 + P_0\int_{x=0}^{x=t}\mathrm{d}H(\underline{r},x)$$

where

$$T(\underline{r},t) = T_0 + P_0H(\underline{r},t)$$

This expression relates the temperature at a position in space at time t, for a time-independent power function.

From the limits of the function of $H(\underline{r},t)$, the limits at infinite time or at the spatial origin can be evaluated.

$$H(\underline{r},t=\infty) = \lim_{t\to\infty}\frac{1}{4\pi K\Delta}\int\frac{erfc\left\{\frac{|\underline{r}-\underline{r}'|}{2\sqrt{Dt}}\right\}}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}' = \frac{1}{4\pi K\Delta}\int\frac{\lim\limits_{t\to\infty}erfc\left\{\frac{|\underline{r}-\underline{r}'|}{2\sqrt{Dt}}\right\}}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}'$$

For small x, the error function $erf\,(x)$ can be expressed as

$$erf\,(x) \simeq \frac{2x}{\sqrt{\pi}} \qquad x \leqslant \sqrt{\pi/2}$$

Then, as t approaches infinity, the complementary error function $erfc\,(0) = 1$, hence

$$H(\underline{r},t=\infty) = \frac{1}{4\pi K\Delta}\int\frac{\lim\limits_{t\to\infty}erfc\left\{\frac{|\underline{r}-\underline{r}'|}{2\sqrt{Dt}}\right\}}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}' = \frac{1}{K\Delta}\int\frac{\mathrm{d}\underline{r}'}{|4\pi(\underline{r}-\underline{r}')|}$$

Hence, the temperature at steady state (e.g. infinite time) can be expressed as

$$T_{ss} = T_0 + P_0 H(\infty) = T_0 + \frac{P_0}{K\Delta} \int \frac{d\underline{r}'}{4\pi|\underline{r} - \underline{r}'|}$$

In the DFC model (Figure 2.5), the objective was to provide a solution associated with a Green's function more representative of the geometry associated with semiconductor devices. For a three-dimensional representation, a parallelepiped is formed in the three physical dimensions with characteristic lengths of a, b, and c. A Green's function is defined at the symmetrical center which represents the one-dimensional solution for a constant heat source at position x within dimension a.

$$G(x, a, \tau) = \frac{1}{2}\left(erf\left(\frac{\frac{a}{2} + x}{2\sqrt{D\tau}}\right) + erf\left(\frac{\frac{a}{2} - x}{2\sqrt{D\tau}}\right)\right)$$

In this case, the Green's function is represented by the error function $erf\,(x)$

$$erf\,(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\} dt$$

Using the constant power pulse form, where the power is outside the integral expression, we can express the temperature at position r as the product of the three Green's functions. Hence a general expression inside the parallelepiped is

$$T(\underline{r}, t) = T_0 + \frac{P_0}{\rho C_p \Delta} \int_0^t G(x, a, \tau) G(y, b, \tau) G(z, c, \tau) d\tau$$

Assuming the peak temperature is at the center of the defect region, we can solve for the geometric center as

$$T(\underline{0}, t) = T_0 + \frac{P_0}{\rho C_p \Delta} \int_0^t G(x = 0, a, \tau) G(y = 0, b, \tau) G(z = 0, c, \tau) d\tau$$

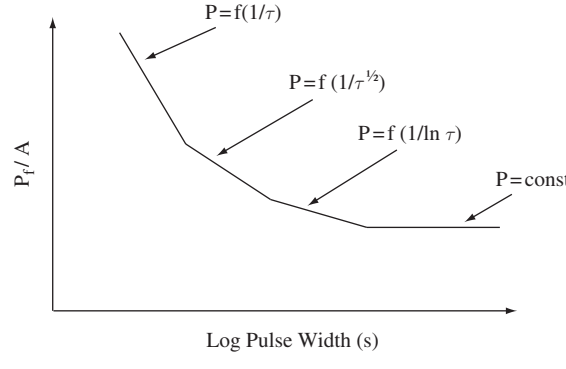

Figure 2.5 Dwyer model

Substituting into the Green's function, the function simplifies to

$$G(x=0,a,\tau)=\frac{1}{2}\left(erf\left(\frac{\frac{a}{2}}{2\sqrt{D\tau}}\right)+erf\left(\frac{\frac{a}{2}}{2\sqrt{D\tau}}\right)\right)=erf\left(\frac{a}{4\sqrt{D\tau}}\right)$$

Hence, at the geometric center, the temperature at time t is

$$T(\underline{0},t)=T_0+\frac{P_0}{\rho C_{\mathrm{p}}\Delta}\int_0^t erf\left(\frac{a}{4\sqrt{D\tau}}\right)erf\left(\frac{b}{4\sqrt{D\tau}}\right)erf\left(\frac{c}{4\sqrt{D\tau}}\right)\mathrm{d}\tau$$

The error functions can be written in dimensionless form by defining the thermal diffusion times. In the DFC model, the three thermal diffusion times were defined as

$$t_a=\frac{a^2}{4\pi D}$$

$$t_b=\frac{b^2}{4\pi D}$$

$$t_c=\frac{c^2}{4\pi D}$$

In this form, the center temperature can be represented as the ratio of the thermal diffusion times to the characteristic time.

$$T(\underline{0},t)=T_0+\frac{P_0}{\rho C_{\mathrm{p}}\Delta}\int_0^t erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_a}{\tau}}\right)erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_b}{\tau}}\right)erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau$$

If the time t is the failure time at which the temperature reaches failure temperatures at the origin, we can express the power to failure as

$$P_{\mathrm{f}}=\frac{\rho C_{\mathrm{p}}\Delta(T(0,t)-T_0)}{\int_0^{t=tf} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_a}{\tau}}\right)erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_b}{\tau}}\right)erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau}$$

To evaluate the limits of the error function, it can be expanded in series form.

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp\{-t^2\}\mathrm{d}t=\frac{2}{\sqrt{\pi}}\int_0^x\left(1-t^2+\frac{t^4}{2!}-\frac{t^6}{3!}+\ldots\right)\mathrm{d}t$$

Completing the integration over the series expansion

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp\{-t^2\}\mathrm{d}t=\frac{2}{\sqrt{\pi}}\left(x-\frac{x^3}{3}+\frac{x^5}{5(2!)}-\frac{x^7}{7(3!)}+\cdots\right)$$

As x is small

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\} \mathrm{d}t \simeq \frac{2x}{\sqrt{\pi}} \qquad x \leqslant \sqrt{\pi}/2$$

and

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\} \mathrm{d}t \simeq 1 \qquad x \geqslant \sqrt{\pi}/2$$

Hence for the terms in the power expression, the error functions are dependent on the relationship between the thermal diffusion coefficients and the characteristic time.

For the three dimensions, assume an ordering of the dimensions where $c < b < a$, leading to the thermal diffusion coefficient ordering of $t_c < t_b < t_a$. Hence the first time regime of interest is the time from $t = 0$ to $t = t_c$. In the temperature integral equation, for this time regime, the three error functions equal unity since the time is shorter than the thermal diffusion time of the three expressions. As a result

$$T(\underline{0}, t) = T_0 + \frac{P_0}{\rho C_\mathrm{p} abc} \int_0^t \mathrm{d}\tau$$

If the failure occurs in this time frame, the power is equal to the power to failure and failure at the peak temperature position is

$$P_\mathrm{f} = \frac{\rho C_\mathrm{p} \Delta (T(0, t = t_c) - T_0)}{t_\mathrm{f}}$$

For times between t_c and t_b the expression for the error function for the z-direction is estimated as the first term in the power series expression, while the x- and y-direction error functions remain at a value of unity. Hence we can express this as

$$P_\mathrm{f} = \frac{\rho C_\mathrm{p} \Delta (T(0,t) - T_0)}{\int_0^{t=t_f} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau} = \frac{\rho C_\mathrm{p} \Delta (T(0,t) - T_0)}{\int_0^{t_c} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau + \int_{t_c}^{t=t_f} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau}$$

As was shown in the previous development

$$P_\mathrm{f} = \frac{\rho C_\mathrm{p} \Delta (T(0,t) - T_0)}{\int_0^{t=t_f} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau} \cong \frac{\rho C_\mathrm{p} \Delta (T(0,t) - T_0)}{\int_0^{t_C} \mathrm{d}\tau + \int_{t_C}^{t=t_f} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau}$$

Substituting the approximation for the error function,

$$\int_{t_c}^{t} erf\left(\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right)\mathrm{d}\tau \simeq \int_{t_c}^{t}\mathrm{d}\tau\frac{2}{\sqrt{\pi}}\left\{\frac{\sqrt{\pi}}{2}\sqrt{\frac{t_c}{\tau}}\right\}$$

$$= \sqrt{t_c}\int_{t_c}^{t}\mathrm{d}\tau\left\{\tau^{-1/2}\right\} = 2\sqrt{t_c}\left[\sqrt{\tau}\right]_{t_c}^{t} = 2\left(\sqrt{t_c\tau} - t_c\right)$$

$$P_{\mathrm{f}} = \frac{ab\sqrt{(\pi K\rho C_{\mathrm{p}})}(T_c - T_0)}{\sqrt{t_{\mathrm{f}}} - \sqrt{t_c}/2}$$

$$P_{\mathrm{f}} = \frac{MC_{\mathrm{p}}(T_c - T_0)}{t_{\mathrm{f}}}$$

$$P_{\mathrm{f}} = \frac{ab\sqrt{(\pi K\rho C_{\mathrm{p}})(T_c - T_0)}}{\sqrt{t_f} - \sqrt{t_c}/2}$$

$$P_{\mathrm{f}} = \frac{4\pi Ka(T_c - T_0)}{\ln(t_{\mathrm{f}}/t_b) + 2 - (c/b)}$$

$$P_{\mathrm{f}} = \frac{2\pi Ka(T_c - T_0)}{\ln\left(\frac{a}{b}\right) + 2 - \frac{c}{2b} - \sqrt{\frac{t_a}{t_{\mathrm{f}}}}}$$

2.2.7 Greve Model

Polysilicon resistor elements are used in circuit networks as resistor elements and fuses. Polysilicon resistors are used for receiver networks in series with MOSFET gates, RC-triggered MOSFET power clamps, and grounded gate NMOS (GGNMOS) networks. For a polysilicon fuse, it is critical to provide a model to understand the conditions for programming current for resistive change or establishing an open circuit.

Greves proposed a polysilicon fuse model [11]. In the fuse structure, a polysilicon structure of thickness d_{p} is placed on a oxide film of thickness d_{ox}. Based on Greves's experimental work, below the critical power P_{crit}, it was observed that the resistance of the polysilicon device did not undergo significant changes in the series resistance. A small decrease in resistance attributed solid phase annealing from the pulse was observed. For pulse power above the critical power condition P_{crit}, a low resistance state was observed in the polysilicon elements. Under this condition, the polysilicon resistance increases as a result of polysilicon material transport prior to an 'open' occur in the film. Greves attributed the low resistance state to filamentary conduction and the transition to a molten state. Greves observed that the estimated fuse temperature was approximately the silicon melting temperature, as well as a model assuming filamentary conduction. Additionally, observation of the molten state was evident. It is postulated that the region of the fuse that achieves the peak temperature is the starting location of

the molten filament. As the temperature increases, current crowding occurs in the filamentary area of the polysilicon film. As the current crowding occurs in the physical region, the size of the filament expands in physical area and distributes in two dimensions, toward the edges of the film and toward the contacts.

Assuming a polysilicon resistor element of thickness d_p and width w on an oxide film of thickness d_{ox}. Assuming a filament of width δ, a molten silicon region is in parallel with a second nonmolten region of width $w - \delta$. The peak temperature across the width is at the symmetric center. The temperature cross section in the polysilicon resistor is such that the temperature between the center and width $\delta/2$ is above the melting temperature (e.g., $T > T_m$ for $\delta/2 < x < (w+\delta)/2$)).

From the heat equation, it can be shown that the relationship between the applied voltage and the filament can be evaluated as

$$V(\delta) = \left[\frac{l^2 g_{ox} T_m}{\sigma d_p d_{ox}} \frac{1 - F(\delta)}{F^* - F(\delta)}\right]$$

$$F(\delta) = \frac{\tanh\{\alpha(\delta - w)/2\}}{\tanh\{\alpha\delta/2\}}$$

and

$$\alpha = \left[\frac{r_{si} g_{ox}}{d_p d_{ox}}\right]^{1/2}$$

2.2.8 Negative Differential Resistance Model

Electrothermal instability can occur in a structure that undergoes a negative resistance state. Assuming a one-dimensional structure, we can calculate a resistor as a summation of incremental resistor elements or integral form

$$R = \sum_{i=1}^{i=n} R_i \Delta x = \int_0^L R(x)\mathrm{d}x$$

From this expression, for a constant film thickness t and width w, a sheet resistance can be defined

$$\rho_{sh} = \frac{\rho}{t_{film}}$$

then

$$R = \int_{x=0}^{x=L} \frac{\rho_{sh}(x, T)}{W}\mathrm{d}x$$

Let the sheet resistance be a function of the position and local temperature T. Taking the derivative of the resistance

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{x=0}^{x=L} \frac{\rho_{\mathrm{sh}}(x,T)}{W}\mathrm{d}x$$

Since the derivative in time is independent of the integration, the derivative and the integral can be interchanged, with

$$\frac{dR}{dt} = \frac{1}{W}\int_{x=0}^{x=L} \frac{\mathrm{d}}{\mathrm{d}t}\rho_{\mathrm{sh}}(x,T)\mathrm{d}x$$

From the total differential

$$d\rho_{\mathrm{sh}}(x,T) = \frac{\partial\rho}{\partial T}\mathrm{d}T + \frac{\partial\rho}{\partial x}\mathrm{d}x$$

and

$$\frac{\mathrm{d}\rho_{\mathrm{sh}}(x,T)}{\mathrm{d}t} = \frac{\partial\rho}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} + \frac{\partial\rho}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}$$

Since $\mathrm{d}x/\mathrm{d}t = 0$, the second term in the expression is zero, leaving the expression as

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{1}{W}\int_{x=0}^{x=L} \frac{\partial}{\partial T}\rho_{\mathrm{sh}}(x,T)\frac{\partial T(x,T)}{\partial t}\mathrm{d}x$$

Negative differential resistance (NDR) instability is present if the following expression is less than zero. The system is metastable at a value of zero.

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{1}{W}\int_{x=0}^{x=L} \frac{\partial}{\partial T}\rho_{\mathrm{sh}}(x,T)\frac{\partial T(x,T)}{\partial t}\mathrm{d}x \leqslant 0$$

From this development, it is also clear that the system is stable if

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{1}{W}\int_{x=0}^{x=L} \frac{\partial}{\partial T}\rho_{\mathrm{sh}}(x,T)\frac{\partial T(x,T)}{\partial t}\mathrm{d}x \geqslant 0$$

Let us assume that a segment of the resistance from $x = L_1$ to $x = L$ satsifies the condition

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{1}{W}\int_{x=L1}^{x=L} \frac{\partial}{\partial T}\rho_{\mathrm{sh}}(x,T)\frac{\partial T(x,T)}{\partial t}\mathrm{d}x \leqslant 0$$

Then, if

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{1}{W}\int_{x=0}^{x=L_1} \frac{\partial}{\partial T}\rho_{\mathrm{sh}}(x,T)\frac{\partial T(x,T)}{\partial t}\mathrm{d}x + \frac{1}{W}\int_{x=L_1}^{x=L} \frac{\partial}{\partial T}\rho_{\mathrm{sh}}(x,T)\frac{\partial T(x,T)}{\partial t}\mathrm{d}x \geqslant 0$$

then the region from resistor from $x=0$ to $x=L_1$ has stabilized the negative differential resistance. If this condition is true for all times, then the first integral expression serves as the resistor ballasting condition for achieving stability.

2.2.9 Ash Model

M. Ash revisited the models proposed by Wunsch and Bell, realizing that the material properties varied significantly with temperature and yet the early models did not assume temperature-dependent coefficients [7]. In the Wunsch–Bell model, the derivation assumes the thermal conductivity and density-specific heat product is a constant. The model assumes a one-dimensional thermal diffusion equation with constant coefficients. The boundary conditions also assume a semi-infinite medium with the junction on a single face of the medium. Ash used the Kirchoff and Boltzmann transformations to solve the nonlinear power to failure solution of the heat equation.

Thermal conductivity, density and specific heat are all temperature dependent in semiconductor materials such as silicon, germanium, and gallium arsenide. Let us assume that the properties are not constant.

The heat equation can be expressed as a one-dimensional temperature-dependent thermal conductivity $K(T)$ and temperature-dependent density – specific heat product,

$$\frac{\partial}{\partial x}\left(\kappa(T)\frac{\partial T}{\partial x}\right) = \rho c_{\mathrm{p}}(T)\frac{\partial T}{\partial t}$$

By transformation of variables, let us define a spatial and time temperature $Q(x,t)$

$$Q(x,t) = \int_{T_0}^{T} \sqrt{K(T')\rho Cp(T')}\mathrm{d}T'$$

and the transformation of the position as

$$\chi(x) = \int_0^x \sqrt{\frac{\rho C_{\mathrm{p}}}{K}}\mathrm{d}x'$$

Ash showed that this leads to a transformation of

$$\ln \varsigma = -C\overline{Q}(\chi,t)$$

$$\overline{Q}(\chi(x,t),t) = Q(x,t)$$

with boundary conditions of

$$\varsigma(\chi, 0) = 1$$

$$\varsigma(\infty, t) = 1$$

and the transformed differential equation as

$$\frac{\partial^2}{\partial^2 \chi}\varsigma = \frac{\partial}{\partial T}\varsigma + \frac{CP_0}{A}\frac{\partial}{\partial \chi}\varsigma$$

where

$$\frac{\frac{\partial}{\partial \chi}\varsigma(0,t)}{\varsigma(0,t)} = \frac{CP_0}{A}$$

In order for the system of equations to be valid, the thermal conductivity and the specific heat – density product has to satisfy a first order ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}T}\ln\sqrt{\rho C_\mathrm{p}/K} = C\sqrt{\rho C_\mathrm{p} K}$$

This first-order differential equation can be expressed as an integral solution

$$\left(\frac{\rho C_\mathrm{p}}{K}\right) = C\left(a_0 + \int_{T_0}^{T} \rho C_\mathrm{p}(T')\mathrm{d}T'\right)$$

A general formulation can be established with an arbitrary function of the form

$$K(T) = K_0 g(T)\exp\left\{-C\sqrt{(\rho C_\mathrm{p})_0 K_0}\int_{T_0}^{T} g(T')\mathrm{d}T'\right\}$$

$$\rho C_\mathrm{p}(T) = (\rho C_\mathrm{p})_0 g(T)\exp\left\{+C\sqrt{(\rho C_\mathrm{p})K}\int_{T_0}^{T} g(T')\mathrm{d}T'\right\}$$

Substituting this into the first-order differential equation,

$$\frac{\mathrm{d}}{\mathrm{d}T}\ln\sqrt{\rho C_\mathrm{p}/K} = C\sqrt{\rho C_\mathrm{p} K}$$

Substituting in the general expression

$$\frac{\mathrm{d}}{\mathrm{d}T}\ln\sqrt{\rho C_\mathrm{p}/K} = \frac{\mathrm{d}}{\mathrm{d}T}\ln\sqrt{\frac{(\rho C_\mathrm{p})_0 g(T)\exp\{-C\sqrt{(\rho Cp)_0}\int_{T_0}^{T} g(T')\mathrm{d}T'\}}{K_0 g(T)\exp\{C\sqrt{(\rho Cp)_0}\int_{T_0}^{T} g(T')\mathrm{d}T'\}}}$$

and

$$C\sqrt{\rho CpK} = C\sqrt{\left\{(\rho C_{\mathrm{p}})_0 g(T)\exp\{-C\sqrt{K_0(\rho C_{\mathrm{p}})_0}\int_{T_0}^{T} g(T')\mathrm{d}T'\}\right\}}$$
$$\sqrt{K_0 g(T)\exp\{C\sqrt{K_0(\rho C_{\mathrm{p}})_0}\int_{T_0}^{T} g(T')\mathrm{d}T'\}}$$

where

$$C\sqrt{\rho C_{\mathrm{p}} K} = Cg(T)\sqrt{K\rho C_{\mathrm{p}}}$$

The function $g(T)$ is arbitrary, but must normalize to $g(T_0) = 1$.

The differential equation can be rearranged such that the left-hand side equals a constant.

$$\frac{1}{\sqrt{\rho C_{\mathrm{p}} K}}\frac{\mathrm{d}}{\mathrm{d}T}\ln\sqrt{\rho C_{\mathrm{p}}/K} = C$$

Functions that demonstrate this relationship will be solutions to the general equation with the variable coefficients. Ash showed that for many materials such as silicon, germanium and gallium arsenide, the relationship is valid over a wide temperature range. The mystery that Ash was concerned with was although the nonlinearity was not addressed, the experimental data fit the Wunsch–Bell model. Part of the reason was that the nature of the materials that were used for the Wunsch–Bell model satisfied the Ash relationships.

2.2.10 Statistical Models

Prediction of ESD results is a function of the statistical variations of the structure as well as the statistical variations of the pulse event [12–15]. Statistics of failure are important in order to predict failure in a real environment.

The statistical variation of a nanostructure device is dependent on the variations of the geometrical dimensions, and doping concentrations. In a semiconductor device, the variations of the geometrical dimensions are dependent on the photolithography, etch process, implant, and diffusion processes. In these processes, random as well as systematic effects exist. There are microscopic and macroscopic variations leading to both 'local' and 'global' changes.

Additionally, the macroscopic global variations are on a higher scale associated with manufacturing process variations. Manufacturing process variations can consist of within wafer die-to-die, wafer-to-wafer, lot-to-lot, and semiconductor fabricator site-to-site. These variations can be caused by process variations within a semiconductor tool, tool-to-tool, as well as incoming wafers. It is also not uncommon that multiple suppliers produce a given semiconductor chip or product by completely different semiconductor processes.

Statistical variations also exist in the simulated ESD testers. The electrical discharge imposed on the semiconductor chips have both voltage and current variations in the electrical discharge applied.

In the early development of semiconductors, it was important for applications to be able to provide good prediction of failure. The relationship between ESD failures and the statistics of failure were investigated by Brown [12], Enlow [13], Alexander [14], and Pierce and Mason [15]. Statistical methods were developed which utilized the physics of failure models combined with variation assumptions in the semiconductor industry.

If we define a variable which is the ratio of the mean failure power $< P_{\mathrm{f}} >$ and the sample standard deviation S_{p}, for N discrete elements

$$\Gamma = \frac{< P_{\mathrm{f}} >}{S_{\mathrm{p}}}$$

where we define

$$S_{\mathrm{p}} = \left[\frac{1}{N-1} \sum_{i=1}^{N} \{ P_{\mathrm{f}_i} - \langle P_{\mathrm{f}} \rangle \}^2 \right]^{1/2}$$

Assuming a lognormal distribution we can express a probability distribution function as

$$f_{\Gamma}(\lambda) = \frac{1}{\sqrt{2\pi}\ \ Sp} \exp\left\{ -1/2 \left[\frac{\ln \lambda - \langle P_f \rangle}{Sp} \right] 2 \right\}$$

Letting λ be a random variable, we can express a cumulative distribution function as

$$\Pr\{\Gamma > \lambda_N\} = \int_{\lambda_N}^{\infty} f_{\Gamma}(\lambda) \mathrm{d}\lambda$$

The proper interpretation of this equation is that N% of the values of variable λ are greater than the parameter λ_N as stated by Pierce and Mason [15]. To use the Pierce–Mason development (Figure 2.6), for a fixed value of device N, we obtain the variable λ_N. Since the mean power to failure is known, the standard deviation is calculated as

$$S_{\mathrm{p}} = \frac{< P_{\mathrm{f}} >}{\lambda_N}$$

As noted by Pierce and Mason, these results imply that, given the value for the variable λ_N, the standard deviation of the threshold distribution is bounded by a linear function of the mean power to failure.

As an application of the statistics, Enlow showed that a linear relationship exists between the logarithm of the standard deviation S_{p} and the logarithm of the mean power to failure $< P_{\mathrm{f}} >$.

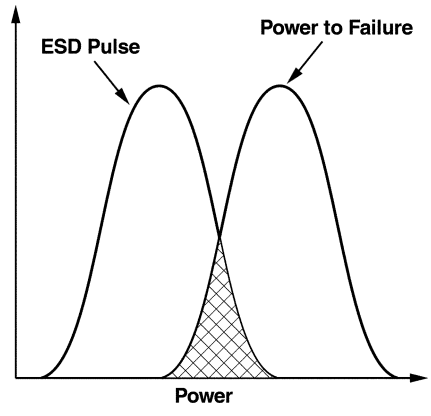

Figure 2.6 Pierce–Mason statistical model

Assuming a Wunsch–Bell failure dependence in the thermal diffusion regime, let us state this by lumping all the physical parameters in the constant K, and representing the relation as

$$P_{\mathrm{f}} = \frac{K\delta}{\sqrt{\tau_{\mathrm{p}}}}$$

Assuming the statistical variations are associated with the processing variations associated with area δ, we denote these as mean variation $<\delta>$ and standard deviation S_δ, and then

$$\langle P_{\mathrm{f}} \rangle = \frac{K}{\sqrt{\tau_{\mathrm{p}}}} \langle \delta \rangle$$

and

$$S_{\mathrm{p}} = \frac{K}{\sqrt{\tau_{\mathrm{p}}}} S_\delta$$

and hence by substitution, we obtain the relationship

$$S_{\mathrm{p}} = \langle P_{\mathrm{f}} \rangle \frac{S_\delta}{\delta}$$

To address the case where the location of the failure is not known, a random variable can be defined. Pierce and Mason defined a new random variable which is the ratio of the analytically estimated mean failure threshold and the measured mean failure threshold.

$$\alpha = \frac{\langle P_{\mathrm{c}} \rangle}{\langle P_{\mathrm{f}} \rangle}$$

A probability distribution function can be defined as a two-parameter distribution where

$$f_{P_f} = f_{P_f}(P_f, \langle P_f \rangle, S_p)$$

Substituting in the definitions from the Pierce–Mason model, a distribution function can be defined such that a device of failure power P_f can survive a pulse of power P_f'

$$\Pr\{P_f > P_{f'}\} = \int_{P_{f'}}^{\infty} f_{P_f}(P_f, \langle P_c \rangle/\alpha, \langle P_c \rangle/\alpha\lambda) \mathrm{d}P_f$$

Statistical methods are of value for building high-reliability systems where ESD failures are unacceptable. Whereas these statistical methods were used in the early 1970s and 1980s to address ESD reliability and prediction, the solutions to addressing ESD concerns were not typically statistical in nature in the 1980s and 1990s. In the 1990s, and recent times, the ESD solution has addressed this by designing products to higher ESD levels to reduce ESD risks. On-chip protection that addresses all the mechanisms by increasing the ESD margins between events and environment has been an effective solution since the 1990s. With the increasing difficulty of building semiconductors, as well as the distribution of their assembly on a global level, controlling all factors is more difficult. As devices are scaled, the margin between the sensitivity of the product and the events will be reduced, leading to a reinstitution of the statistical methods and product risk assessments.

The next chapter will focus on semiconductor devices, dealing with only certain areas and devices of interest. The devices include diodes, resistors, bipolar transistors, silicon-controlled rectifiers, and MOSFET high-current, high-voltage and high-temperature effects. The chapter will address leakage phenomenon from thermal models, trap-to-band models, band-to-band tunneling models which are important for understanding leakage and becomes more critical as oxides are scaled to smaller dimensions. For resistor elements, the focus will be on velocity saturation effects. For bipolar transistors, we will discuss limitations such as the Johnson Limit, Kirk effect and basic models. For the silicon-controlled rectifier, the focus will be on the criteria for triggering for different cases, and the general tetrode model. Holding current formulations will also be discussed. For MOSFETs, the focus will be on MOSFET current constriction, avalanche phenomenon, parasitic bipolar, and MOSFET gate-induced drain leakage phenomena.

PROBLEMS

2.1 Assuming a Wunsch–Bell form of the power to failure in the thermal diffusion regime, an ideal current source, an input power of an element with a breakdown voltage and a series resistance

$$P_f = \frac{K}{\sqrt{t_p}}$$

and

$$\int_0^t \frac{P(\lambda)}{2\sqrt{t-\lambda}}\mathrm{d}\lambda \geqslant K$$

$$I(t) = I_0 e^{-t/\tau}$$

$$P(t) = I(t)V_{BD} + I^2(t)R_\mathrm{s}$$

derive the following inequality

$$I_0 V_\mathrm{BD}\sqrt{t}D\left(\sqrt{t/\tau}\right) + {I_0}^2 R_\mathrm{s}\sqrt{\tau/2}D\left(\sqrt{2t/\tau}\right) \geqslant K$$

where $D(x)$ is the Dawson integral

$$D(x) = e^{-x^2}\int_0^x e^{y^2}\mathrm{d}y$$

2.2 Given the power to failure from the Wunsch–Bell thermal model, and the maximum power expression of the Johnson limit, derive the dimensionless group of the power to failure normalized to the maximum power of the Johnson Limit. What other dimensionless groups and time constants are contained within the expression? What physical processes are competing?

REFERENCES

1. J. Mathews and R. L. Walker. *Mathematical Methods of Physics*, W. A. Benjamin., 1964.
2. P. M. Morse and H. Feshbach. *Methods of Theoretical Physics*, McGraw-Hill, 1953.
3. H. Weinberger. *A First Course in Partial Differential Equations*, Wiley 1965.
4. D. M. Tasca. Pulse power failure modes in semiconductors. *IEEE Transactions on Nuclear Science* 1970; **NS-17** (6): 346–372.
5. D. C. Wunsch and R. R. Bell. Determination of threshold voltage levels of semiconductor diodes and transistors due to pulsed voltages. *IEEE Transactions on Nuclear Science* 1968; **NS-15** (6): 244–259.
6. J. S. Smith and W. R. Littau. Prediction of thin-film resistor burn-out. *Proceedings of EOS/ESD Symposium*, 1981; 192–197.
7. M. Ash. Semiconductor junction non-linear failure power thresholds: Wunsch–Bell revisited. *EOS/ESD Symposium*, 1983; 122-127.
8. V. I. Arkihpov, E. R. Astvatsaturyan, V. I. Godovosyn and A. I. Rudenko. *International Journal of Electronics* **55**: 1983; 395.
9. V. A. Vlasov and V. F. Sinkevitch. *Elektronnaya Technika*, **4**: 1971; 68–75.
10. V. M. Dwyer, A. J. Franklin and D. S. Campbell. Thermal failure in semiconductor devices. *Solid State*, 1989; 553–560.
11. D. W. Greve. Programming mechanism of polysilicon fuse links. *International Electron Device Meeting (IEDM) Technical Digest*, 1981; 70-74.

12. W. D. Brown. Semiconductor device degradation by high amplitude current pulses. *IEEE Transactions On Nuclear Science* 1972; **NS-19**: 68–75.
13. D. R. Alexander and E. W. Enlow. Predicting lower bounds on failure power distributions of silicon *npn* transistors. *IEEE Transactions on Nuclear Science* 1981; **NS-28** (6): 145–150.
14. E. N. Enlow. Determining an emitter-base failure threshold density of *npn* transistors. *Proceedings of the EOS/ESD Symposium*, 1981; .
15. D. Pierce and R. Mason. A probabilistic estimator for bounding transistor emitter-base junction transient-induced failures. *Proceedings of the EOS/ESD Symposium*, 1982; 82–90.

3 Semiconductor Devices and ESD

In this chapter, we discuss basic semiconductor components and ESD phenomena. The chapter will first address semiconductor equations. This is followed by analysis of diodes, bipolar transistors, silicon-controlled rectifiers, resistors and MOSFETs. Since the literature is filled with the discussion on semiconductors, we will address the high-current effects, and high-electric-field issues that are concerns in ESD analysis. The chapter will focus on thermal effects as well.

3.1 DEVICE PHYSICS

For the electrothermal physics of semiconductors, the fundamental equations of semiconductors and corresponding constitutive relationships must be solved in order to evaluate the electron and hole concentrations, the current densities and local temperature [1–7]. A fully coupled methodology solves for the carrier density properties and temperature in the fundamental set of equations [1]. In this fashion, the electrical potential, carrier densities and temperature can be solved for self-consistently. In the formulation below, one temperature of the system is assumed locally. This assumption inherently requires that the electron, hole and lattice temperature are equal ($T = T_e = T_h = T_{lattice}$).

The fundamental equations in an electroquasistatic (EQS) formulation includes the Poisson equation for semiconductors,

$$\nabla \bullet (\varepsilon \nabla \psi) = -q(p - n + N_D + N_A)$$

where ε is the dielectric constant, p is the hole concentration, n is the electron concentration, N_D is the donor concentration, and N_A is the acceptor concentration. The

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

current continuity equations for the electron and the hole concentration can be expressed as

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla \bullet J_n = -R$$
$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla \bullet J_p = -R$$

The conservation of energy (or heat energy continuity) can be expressed as

$$C\frac{\partial T}{\partial t} + \nabla \bullet J_Q = H$$

From this form, the constitutive relationships for electron current, hole current and heat energy flux can be established. The constitutive relationship for electrons contains an electron drift term, and an electron diffusion term when the temperature of the system is constant. An additional current drive term is present when there is a gradient in the temperature.

$$J_n = q\mu_n n\left(E + \frac{kT/q}{n}\nabla n + \frac{5}{2T}\frac{kT}{q}\nabla T\right)$$

The hole constitutive relationship of drift–diffusion is

$$J_p = q\mu_p p\left(E - \frac{kT/q}{p}\nabla p - \frac{5}{2T}\frac{kT}{q}\nabla T\right)$$

and heat energy flow is

$$J_Q = -K(T)\nabla T$$

where the thermal conductivity, K is temperature dependent.

In the heat equation, the heat generation in the medium is a function of the Joule heating, the Thomson heat and the recombination heat generation.

The Joule heating can be expressed as the sum of the electron Joule heating and the hole Joule heating terms

$$E = \frac{{J_n}^2}{\sigma_n} + \frac{{J_p}^2}{\sigma_p}$$

The Thomson heat generation term is associated with the thermoelectric power for electrons and holes

$$E = T\{\nabla P_n \bullet J_n - \nabla P_p \bullet J_p\}$$

The recombination heat energy can be expressed as

$$E = qR[\phi_p - \phi_n + T(P_p + P_n)]$$

3.1.1 Nonisothermal Simulation

For very short ESD pulse widths and very fast rise times, the assumption that the electron temperature, hole temperature and lattice temperature are equal may not be valid in the high-current, high-field regions [1,6]. For a general analysis, the constraint of electron temperature, hole temperature and lattice temperature are not equal. The fundamental equations in an EQS formulation includes the Poisson equation for semiconductors

$$\nabla \bullet (\varepsilon \nabla \psi) = -q(p - n + N_{\mathrm{D}} + N_{\mathrm{A}})$$

where ε is the dielectric constant, p is the hole concentration, n is the electron concentration, N_{D} is the donor concentration, and N_{A} is the acceptor concentration. The current continuity equations for the electron and the hole concentration can be expressed in a steady state mode as

$$-\frac{1}{q}\nabla \bullet J_n = G_n - R_n$$
$$+\frac{1}{q}\nabla \bullet J_p = G_p - R_p$$

The thermal conduction equation can be expressed as

$$\nabla \bullet (K(T_L))\nabla T_L = J \bullet E + E_{\mathrm{g}}(R - G)$$

and the carrier energy continuity equations

$$\nabla \bullet S_n = J_n \bullet E - B_n(\delta Te')n$$
$$\nabla \bullet S_p = J_p \bullet E - B_p(\delta T_h')p$$

From this form, three energy balance equations are created where the first equation is associated with the heat transport conduction in the lattice. The thermal conductivity is a function of the lattice temperature and the gradient of the lattice temperature serves as a drive term equal to the Joule heating and the recombination–generation term of the carriers. The carrier energy equations are a function of the electron temperature T_{e}, and the hole temperature T_{h}. The carrier energy of the electrons and holes are related to the energy which is transferred from the lattice to the carries through the electric field, and the energy is lost through energy transfer from the carrier to the lattice via collision.

From this form, the constitutive relationships for electron and hole current can be derived. The constitutive relationship for electrons contains an electron drift term, and

an electron diffusion term; an additional current drive term is present when there is a gradient in the lattice temperature.

$$J_n = -q\mu_n nE + qD_n\nabla n + qnD_n{}^T\nabla T$$
$$J_p = -q\mu_p pE - qD_p\nabla p + qpD_p{}^T\nabla T$$

In this representation, the temperature is the lattice temperature. On a microscopic level, this term can be interpreted as the local electron temperature and local hole temperature

$$J_n = -q\mu_n nE + qD_n\nabla n + qnD_n{}^T\nabla\delta T'_{\mathrm{e}}$$
$$J_p = -q\mu_p pE - qD_p\nabla p + qpD_p{}^T\nabla\delta T'_{\mathrm{h}}$$

where we define a differential hole and electron temperature as the deviation from the lattice temperature

$$\delta T_{\mathrm{e}} = |T_L - T_{\mathrm{e}}|$$

and

$$\delta T_{\mathrm{h}} = |T_L - T_{\mathrm{h}}|$$

The electron and hole concentrations can then be expressed as a function of the quasi-Fermi potential and the respective temperatures as

$$n = n_i\Phi_n \exp\{q\psi/K_B T_{\mathrm{e}}\}$$
$$p = n_i\Phi_p \exp\{-q\psi/K_B T_{\mathrm{h}}\}$$

and the potentials can be expressed as a function of the quasi-Fermi levels as is used in isothermal conditions.

$$\Phi_n = \exp\{-q\varphi_n/K_B T_{\mathrm{e}}\}$$
$$\Phi_p = \exp\{q\varphi_p/K_B T_{\mathrm{h}}\}$$

From the carrier energy continuity equations, the energy loss rate can be expressed as a function of the electron energy relaxation time, and hole energy relaxation time.

$$B_n(\delta T'_{\mathrm{e}}) = \frac{3}{2}\frac{K_B T'_{\mathrm{e}}}{\tau_{\mathrm{we}}}$$
$$B_p(\delta T'_{\mathrm{h}}) = \frac{3}{2}\frac{K_B T'_{\mathrm{h}}}{\tau_{\mathrm{wh}}}$$

To evaluate the ESD robustness of a technology, it is important to understand semiconductor device elements which are used in integrated electronics. To cover all

aspects of the devices in detail would be too exhaustive. In this section, we will focus on some fundamentals, nonlinear behavior, high-electric-field response, and self-heating effects.

3.2 DIODES

Diodes have been a very important electrostatic discharge (ESD) element in the past, as well as remaining an important ESD element in future technologies [8–16]. Diodes have the advantage of low capacitance, and are scalable. Diodes have been able to evolve with the development of NMOS, CMOS, and BiCMOS technology. Diode designs have migrated from CMOS to silicon-on-insulator (SOI) and radio-frequency applications with continued usage in the future [17–21]. As opposed to MOSFET-based ESD networks, diodes do not have gate oxide films sensitive to dielectric failure, from human body model (HBM) to charged device model (CDM) failures. As a result, it is important to fully understand diode process sensitivities, design, and operation. In this chapter, we will discuss the forward-bias and reverse-bias condition. The forward-bias condition is important to understand the ability to use the diode as a current discharge element in a forward-bias mode of operation. The reverse-bias condition is important to understand the leakage mechanisms which have a role in the initial bias condition and failure mechanisms.

Figure 3.1 is an example of an ESD double-diode network. This uses a $p+/n$-well diode for positive ESD pulses and an n-well-to-substrate diode for negative ESD pulses.

3.2.1 Diode Equation

The diode current equation can be obtained by evaluation of the total current through a p–n metallurgical junction. The current flowing through the diode is expressed as a

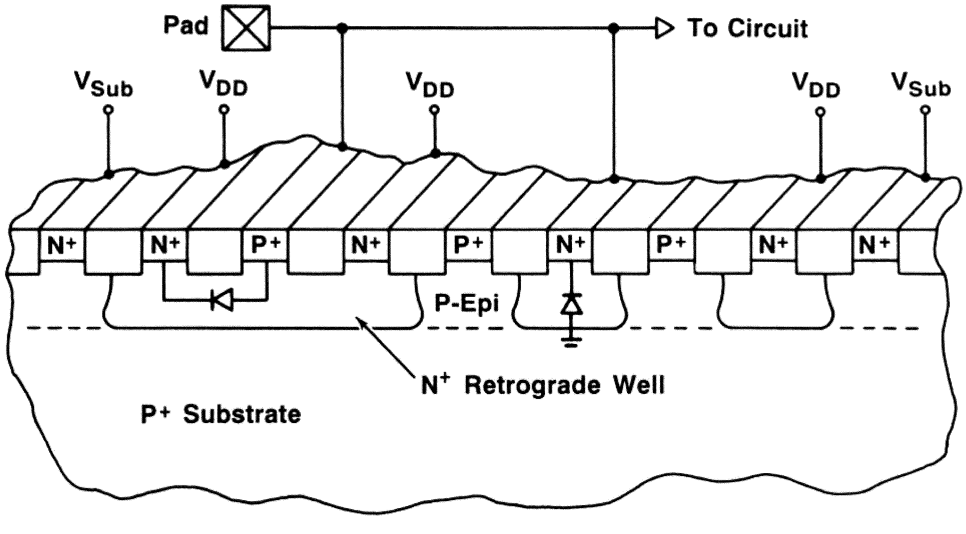

Figure 3.1 Diode-based ESD network

function of the excess minority electron and hole populations at the p–n metallurgical junction.

$$I = qA\left\{\frac{D_p}{L_p}\Delta p_n + \frac{D_n}{L_n}\Delta n_p\right\}$$

where

$$\Delta p_n = p'_n - p_{n0}$$
$$\Delta n_p = n'_p - n_{p0}$$

Under high-level injection [21,22], the excess minority-carrier populations can be expressed as

$$\Delta p_n = \frac{p_n\left(e^{Vj/kT} - 1\right)\left(1 + \dfrac{n_n}{p_p}e^{q(Vj-\phi_i)/kT}\right)}{1 - e^{2q(Vj-\phi_i)/kT}}$$

$$\Delta n_p = \frac{n_p\left(e^{Vj/kT} - 1\right)\left(1 + \dfrac{p_p}{n_n}e^{q(Vj-\phi_i)/kT}\right)}{\left(1 - e^{2q(Vj-\phi_i)/kT}\right)}$$

From this expression, the diode current equation in high-level injection can be expressed as

$$I = I_s\frac{\left(e^{Vj/kT} - 1\right)\left(1 + \eta e^{q(Vj-\phi_i)/kT}\right)}{1 - e^{2q(Vj-\phi_i)/kT}}$$

where

$$I_s = qA\left\{\frac{D_p}{L_p}p_{n0} + \frac{D_n}{L_n}n_{p0}\right\}$$

and for a $p+/n$ diode

$$\eta = \frac{N_a}{N_d}$$

and for an $n+/p$ diode

$$\eta = \frac{N_d}{N_a}$$

This expression reduces to the ideal diode case when the excess minority-carrier injection is small compared with the doping concentrations. The ideal diode model is expressed as

$$I = I_s\left\{\exp\left(\frac{qVj}{kT}\right) - 1\right\}$$

where I is the diode current, and I_s is the reverse saturation current.

In these developments, it assumed that the anode and cathode resistance are negligible. During the ESD event, the anode resistance and the cathode resistance play a key role in the operation of the diode as an ESD element. In a $p+/n$-well diode structure, the anode resistance is a function of the contact resistance, silicide film, and the $p+$ diode implant resistance. The design of the contacts, and the silicide film play a key role in the forward-bias operation and the current density distribution of the diode structure. The cathode resistance (e.g. the n-well region) is a function of the implant profile, the dose, and width of the physical well structure. Figure 3.2 is an example of the current-voltage (I–V) characteristic of a diode structure at low and high current.

3.2.1.1 Diode resistance

Diode resistance is critical in the understanding of the operation of a diode element during electrostatic discharge events.

On the anode, the resistance can be expressed as

$$R_{\text{anode}} = R_{\text{M}} + R_{\text{C}} + (R_{\text{Sal}})_{\text{eff}} + (R_{\text{d}})_{\text{eff}}$$

where the anode resistance is the resistance of the interconnect, the contact, the salicide film, and the anode implant. For the silicide film, the material, area coverage and design of the silicide formed on the junction can alter the effectiveness of the silicide as well as the diffusion area on the ESD protection.

$$R_{\text{cathode}} = R_{\text{M}} + R_{\text{C}} + (R_{\text{Sal}})_{\text{eff}} + (R_{\text{d}})_{\text{eff}} + R_{\text{well}}$$

Equivalently, for the cathode, the cathode resistance is the sum of the resistances in series as well as the well resistance. For the cathode, typically the well resistance is the largest component of resistance which influences the ESD results.

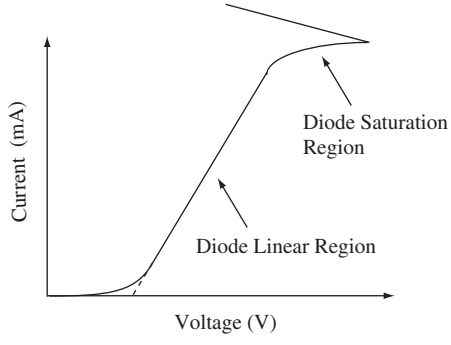

Figure 3.2 Diode current–voltage (I–V) characteristic

From the diode equation, we can express the voltage across the diode as the sum of the voltage drops across the series resistance and the voltage drop across the metallurgical junction

$$I_{\mathrm{D}} = I_{\mathrm{s}}\left\{\exp\left(\frac{qV_{\mathrm{D}} - \sum I_{\mathrm{D}}R_i}{kT}\right) - 1\right\}$$

where summation is over the voltage drops across the series resistances.

The resistance can also be obtained from the derivative of the voltage with respect to current [22].

$$\frac{\mathrm{d}V_{\mathrm{D}}}{\mathrm{d}I_{\mathrm{D}}} = \frac{\mathrm{d}(V_{\mathrm{J}} + V_{\mathrm{R}})}{\mathrm{d}I_{\mathrm{D}}} = \frac{\mathrm{d}V_{\mathrm{J}}}{\mathrm{d}I_{\mathrm{D}}} + \mathrm{R_d}$$

Solving for the resistance

$$R_{\mathrm{d}} = \frac{\mathrm{d}V_{\mathrm{D}}}{\mathrm{d}I_{\mathrm{D}}} - \frac{\mathrm{d}V_{\mathrm{J}}}{\mathrm{d}I_{\mathrm{D}}}$$

From the high-level injection relationship

$$I_{\mathrm{D}} = I_{\mathrm{s}}\exp\left(\frac{qV_{\mathrm{J}}}{2kT}\right)$$

and

$$\frac{\mathrm{d}V_{\mathrm{J}}}{\mathrm{d}I_{\mathrm{D}}} = \frac{2kT}{q}\frac{1}{I_{\mathrm{D}}}$$

Substituting these expression, we can solve for the diode series resistance term as

$$R_{\mathrm{d}} = \frac{\mathrm{d}V_{\mathrm{D}}}{\mathrm{d}I_{\mathrm{D}}} - \frac{2kT}{q}\frac{1}{I_{\mathrm{D}}}$$

3.2.1.2 Self-heating

Diodes undergo both high-level injection and self-heating during an ESD event. This causes the temperature to increase, modulating the diode response. By modification of the diode high-level injection equation during an ESD event, we can calculate the current. In this development, it is assumed that the heating is expressed as the change the junction temperature with time, and that the minority-carrier diffusion is significantly faster than

the rise in temperature. The diode current equation in high-level injection can be expressed as

$$I = I_s \frac{\left(e^{Vj/kT} - 1\right)\left(1 + \eta e^{q(Vj-\phi_i)/kT}\right)}{1 - e^{2q(Vj-\phi_i)/kT}}$$

where

$$I_s = qA\left\{\frac{D_p}{L_p}p_{n0} + \frac{D_n}{L_n}n_{p0}\right\}$$

and for a $p+/n$ diode

$$\eta = \frac{N_a}{N_d}$$

and for a $n+/p$ diode

$$\eta = \frac{N_d}{N_a}$$

Let the junction temperature rise be equal to the product of the thermal impedance and the input power.

$$T_J = \Theta_{TH} P_D$$

The power consumption can be estimated as the power consumption in the diode region,

$$P_D = I_D V_D$$

Hence,

$$T_J = \Theta_{TH}(I_D V_D)$$

where the solution for the diode equation due to high-level injection and self-heating is solved for by correcting the temperature term by the temperature increase. Physical models for the thermal resistance are obtained from the evaluation of the substrate thermal resistance for the diode structure.

Figure 3.3 is an atomic force microscope mapping of a shallow trench isolation (STI) $p+/n$-well diode. The insulating films were removed to observe the silicon interface. Due to the self heating, the insulating region between the $p+$ and $n+$ regions melted, leaving behind molten silicon region between the two diffusion regions.

Figure 3.4 shows an atomic force microscope mapping of a $p+$ diode region in the contact region. The self-heating in the diode structure was significant, leading to the transformation of the silicide region, and melting of the silicon on the surface of the diode structure.

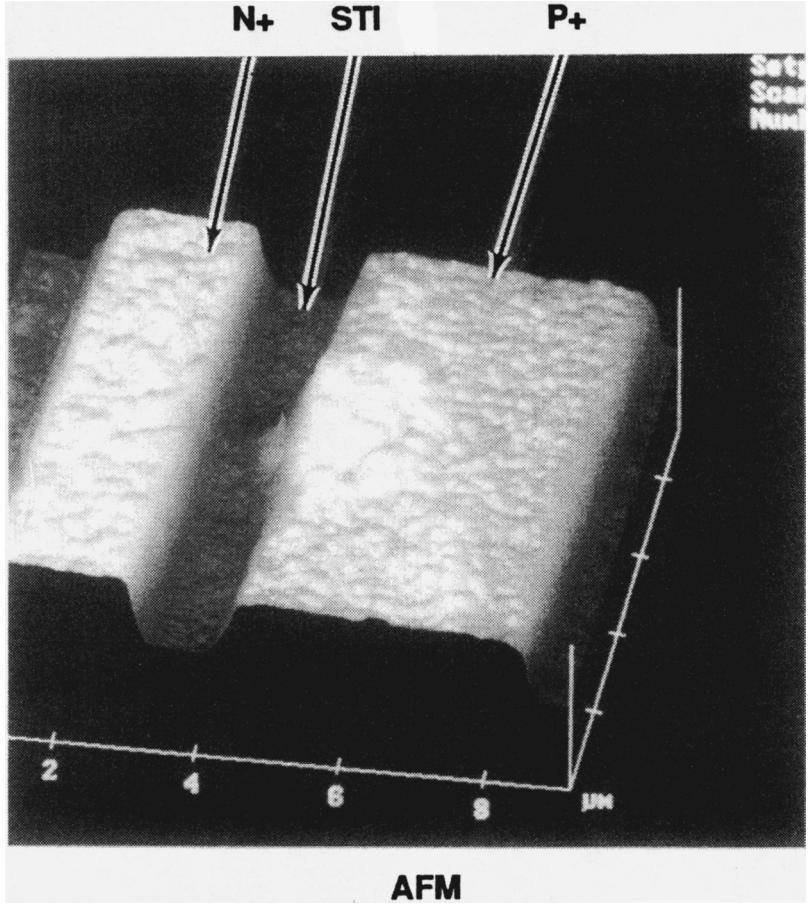

Figure 3.3 Atomic force microscope (AFM) image of diode ESD failure highlighting the p^+/n^+ and shallow trench isolation region

3.2.2 Recombination and Generation Mechanisms

3.2.2.1 Shockley–read–hall model

Understanding of leakage mechanisms in semiconductor devices is important for ESD because leakage is used to determine ESD failure. In the analysis of wafer level ESD testing, or product level, typically only the terminal electrode information is available. The onset of ESD failure is usually determined by semiconductor device electrical parametric shifts. These shifts manifest themselves in changes of the leakage current, resistance shifts, or turn-on voltages.

At low electric fields, leakage current can be explained by the Shockley–Read–Hall (SRH) model [23–25]. The SRH model, in its simplest form, assumes a single trap state energy in the silicon bandgap. These trap states are spatially localized forming independent trap states for carrier capture and of a uniform density. For conduction

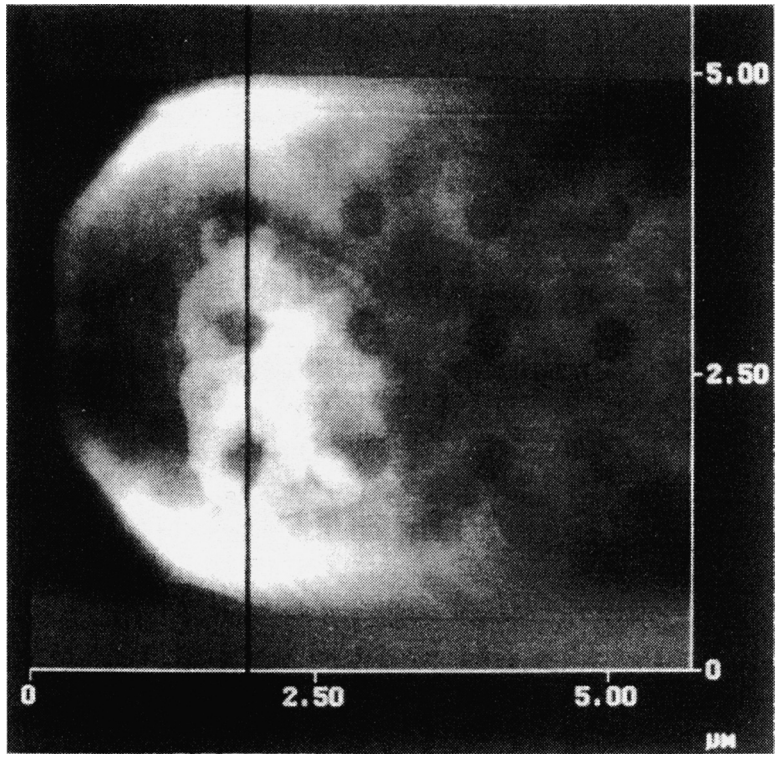

Figure 3.4 Atomic force microscope (AFM) image of diode ESD failure highlighting the p^+ diode implant device surface

band-to-trap transitions, there exists two processes which are inverse processes. The first process is electron capture, and its inverse process is electron emission. For the valence band-to-trap state transitions, the same is true for holes. The capture rate for electrons is

$$r_1 = n[N_t(1 - f(E_t))]v_{th}\sigma_n$$

where n is the number of electrons, N_t is the trap density, $f(E_t)$ is the Fermi–Dirac distribution, v_{th} is the thermal velocity, and the electron capture cross-section is σ. The emission rate is

$$r_2 = n[N_t(f(E_t))]e_n$$

For holes, the capture rate is

$$r_3 = p[N_t(f(E_t))]v_{th}\sigma_p$$

and the hole emission rate is

$$r_4 = [N_t(1 - f(E_t))]e_p$$

These four processes are inherently coupled through the carrier density in the conduction and valence bands. When the system is in equilibrium, the electron capture process is equal to the electron emission process under the thermodynamic law of detailed balance. When the system is not in equilibrium, the capture and emission rates do not balance, leading to a net generation rate.

Let the net generation rate be equal to the difference between the emission rate and the capture rate of the electrons. This is also equal to the difference between the hole emission and capture rate. In order to solve for the generation rate we must eliminate the Fermi–Dirac probability distribution function and substitution into the generation rate formula. The SRH generation rate for a single trap population can be expressed as

$$U = \frac{\left(pn - n_i^2\right)}{\tau_n(p + p_1) + \tau_p(n + n_1)}$$

where

$$\tau_n = \frac{1}{N_t v_{th} \sigma_n}$$

and

$$\tau_p = \frac{1}{N_t v_{th} \sigma_p}$$

and

$$n_1 = n_i \exp\left(\frac{E_t - E_i}{kT}\right)$$

$$p_1 = n_i \exp\left(\frac{E_i - E_t}{kT}\right)$$

3.2.2.2 High-level injection and recombination time constants

Understanding of generation and recombination mechanisms in heavily doped semiconductors are important in ESD and latchup. Generation and recombination mechanisms in heavily doped *n*-wells, *p*-wells, sub-collectors, buried layers, and other device regions is important because of the role these regions play in reduction of minority-carrier injection into substrates and the role they play as high current instabilities.

Recombination and generation physics is a function of the probability of interaction with a carrier population [23,26–28]. We can in general expand a recombination function where recombination is a function of the electron and hole population where the rate is a coefficient. Let

$$R(T) = a_{00}(T) + a_{10}n + b_{01}p + a_{11}np + b_{11}np + a_{21}n^2p + b_{12}np^2 + \ldots$$

In this expansion, the first three terms are unphysical and hence the coefficients can be set to zero, or

$$R(T) = a_{11}np + b_{11}np + a_{21}n^2p + b_{12}np^2 + \ldots$$

According to the principle of detailed balance, each generation term will balance out a corresponding recombination term when the system is in equilibrium. In equilibrium

$$R(T) = a_{11}n_0p_0 + b_{11}n_0p_0 + a_{21}n_0^2p_0 + b_{12}n_0p_0^2 + \ldots = G(T)$$

hence the equilibrium generation term is

$$G_0(T) = (a_{11} + b_{11})n_0p_0 + a_{21}n_0^2p_0 + b_{12}n_0p_0^2 + \ldots$$

Given a system out of equilibrium, where the electron and hole population are perturbed,

$$\frac{Dp'}{Dt} = \frac{\partial p'}{\partial t} + \nabla \bullet Jh = G(T) - R(T)$$

From this form, we can substitute in the electron and hole population in the form

$$n = n_0 + n'$$

and

$$p = p_0 + p'$$

then we can express

$$\frac{Dp'}{Dt} = G_0(T) - R(T) = (a_{11} + b_{11})n_0p_0 + a_{21}n_0^2p_0 + b_{12}n_0p_0^2 - (a_{11} + b_{11})$$
$$(n_0 + n')(p_0 + p') + a_{21}(n_0 + n')^2(p_0 + p') + b_{12}(n_0 + n')(p_0 + p')^2 + \cdots$$

These terms can be sorted into the coefficients and the order of the minority-carrier injection. The first term is the linear first-order term, and the second term is the high-level injection nonlinear term.

$$\frac{\mathrm{d}p'}{\mathrm{d}t} = -\left[\frac{1}{\tau_1} + \frac{1}{\tau_2}\right]p' - \left[\frac{1}{\tau_3}\right]p'^2 - \left[\frac{1}{\tau_4}\right]p'^3$$

where

$$\frac{1}{\tau_1} = (a_{11} + b_{11})(n_0 + p_0)$$
$$\frac{1}{\tau_2} = a_{21}n_0^2 + b_{12}p_0^2 + (a_{21} + b_{12})(2n_0p_0)$$
$$\frac{1}{\tau_3} = a_{21}(2n_0 + p_0) + b_{12}(2p_0 + n_0) + (a_{11} + b_{11})$$
$$\frac{1}{\tau_4} = (a_{21} + b_{21})$$

In this form, the terms are separated by the order of the injection phenomenon. The first term is the term associated with low-level injection phenomenon. The second and third terms are associated with high-level injection. In high-current ESD events, the low-level injection assumption is not always valid.

The first time constant τ_1 is the time constant associated with Shockley–Read–Hall recombination. In this form, it can be seen that the SRH time constant is linearly proportional to the electron and hole population.

The second time constant τ_2 is the Auger recombination time constant. This second time constant is part of the linear response of a semiconductor which is proportional to the square of the doping concentration. The Auger recombination time constant is important at high doping concentrations typically seen in heavily doped retrograde wells, buried layers, and in sub-collectors. This is important for ESD phenomenon and latchup. At high doping concentrations, the time constant τ_2 dominates over the first time constant τ_1.

Hence the linear response of the system (e.g. the low-level injection assumption) is

$$\frac{\mathrm{d}p'}{\mathrm{d}t} = -\left[\frac{1}{\tau_1} + \frac{1}{\tau_2}\right]p'$$

For the nonlinear response where we let τ be the linear time constant

$$\frac{\mathrm{d}p'}{\mathrm{d}t} = -\left[\frac{1}{\tau}\right]p' - \left[\frac{1}{\tau_3}\right]p'^2$$

where

$$\frac{1}{\tau} = \left[\frac{1}{\tau_1} + \frac{1}{\tau_2}\right]$$

Assuming an ESD step pulse at time $t = 0$, an initial condition of excess minority carriers are generated. Separating the variables in the nonlinear ordinary differential equation, with constant coefficients

$$\int_{p'(t=0)}^{p'(t)} \frac{\mathrm{d}p'}{\left[p' + \dfrac{\tau}{\tau_3}p'^2\right]} = \int_{t'=0}^{t'=t} \frac{\mathrm{d}t}{\tau}$$

The solution to the equation, is then

$$p'(t) = \frac{p'(t_0)}{\left[\left(1 + \left(\frac{\tau}{\tau_3}\right)p'(t_0)\tau\right)\exp\left\{\frac{t - t_0}{\tau}\right\} - \left(\frac{\tau}{\tau_3}\right)p'(t_0)\right]}$$

3.2.2.3 Electric-field-enhanced leakage phenomena

In semiconductors, high internal electric fields can lead to leakage phenomena which does not have the current and voltage characteristics of the (SRH) model [26–34]. The SRH model assumes the internal electric field does not influence the capture and emission cross-section. At high internal electric fields, this may not be a valid assumption. The internal electric fields are a function of the doping profiles and the applied electric fields. In an ESD event, motion of the dopants after an ESD impulse can lead to dopant profiles which are more abrupt in the metallurgical junction region. For example, bipolar emitter–base junctions are optimized to avoid tunneling current. Small shifts in the dopant profiles in abrupt or hyper-abrupt metallurgical junctions can lead to ESD failure mechanisms. Additionally, polysilicon films have leakage characteristics which may also have non-SRH characteristics.

The SRH model can be modified to include this phenomenon, by assuming an electric-field-enhanced electron and hole emission rate [29,30],

$$\sigma_n(E) = \sigma_{n0}\aleph_n(E)$$
$$\sigma_p(E) = \sigma_{p0}\aleph_p(E)$$

where the cross-sections are the product of a zero electric field term, and the electric field enhancement factor. In this fashion, high-electric-field phenomenon can be integrated into the SRH model statistics.

$$U = \frac{(pn - n_i^2)}{\tau_n(p + p_1) + \tau_p(n + n_1)}$$

where

$$\tau_n = \frac{1}{N_t v_{th}\sigma_n(E)} = \frac{\tau_{n0}}{\aleph_n(E)}$$

and

$$\tau_p = \frac{1}{N_t v_{th}\sigma_p(E)} = \frac{\tau_{p0}}{\aleph_p(E)}$$

The electric field enhancement factor can be separated into classical and quantum terms.

$$\aleph_n(E) = (\aleph_n(E))_{\mathrm{CL}} + (\aleph_n(E))_{\mathrm{QM}}$$

These enhancement variables in general can be both electric field and thermal dependent.

$$\aleph_n(E,T) = (\aleph_n(E,T))_{\mathrm{CL}} + (\aleph_n(E,T))_{\mathrm{QM}}$$

3.2.2.4 Frenkel–poole effect

From the electric-field-enhanced leakage model

$$U = \frac{(pn - n_i^2)}{\tau_n(p + p_1) + \tau_p(n + n_1)}$$

where

$$\tau_n = \frac{1}{N_t v_{th} \sigma_n(E)} = \frac{\tau_{n0}}{\aleph_n(E)}$$

and

$$\tau_p = \frac{1}{N_t v_{th} \sigma_p(E)} = \frac{\tau_{p0}}{\aleph_p(E)}$$

The electric field enhancement factor can be separated into classical and quantum terms.

At intermediate electric fields, the classical models that include electric field enhancement are the one- and three-dimensional Frenkel–Poole models [31]. The Frenkel–Poole model addresses the influence of the electric field on the trap state barrier height. As the barrier height is modulated by the electric field the emission probability increases, leading to an enhancement of the emission rate, and an increase in the leakage. This model is evident in heavily doped p–n junctions, gated p–n junctions, and polysilicon resistor films. Woo and Plummer applied it to heavily doped emitter–base junction leakage to explain voltage dependent leakage. Voldman applied this model for the explanation of the MOSFET gate-induced drain leakage (GIDL) mechanism in p-channel MOSFETs, and substrate plate trench DRAM capacitors in the low-voltage regime.

In the Frenke–Poole model, the change in the barrier height enhances the probability of the emission probability of a carrier escaping from a trap state. In this model, the analysis is a one-dimensional model. For the Frenkel–Poole effect, the enhancement factor for electron emission for donor-type states and hole emission for acceptor states can be expressed as

$$\aleph_{n\mathrm{D}} = \exp\left[\frac{\Delta\psi}{kT}\right]$$

$$\aleph_{p\mathrm{A}} = \exp\left[\frac{\Delta\psi}{kT}\right]$$

where

$$\Delta\psi = \sqrt{\frac{qE_{ext}}{\varepsilon(kT)}}$$

This model was extended for three-dimensional emission effects by Hartke, by integration around the entire trap in three dimensions [32].

$$\aleph = \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} d\theta \sin\theta \aleph(\theta)$$

Since the electric field is uniform, the emission rate is a function of the angle of emission, where the emission rate is expressed as

$$\aleph_{3D} = \frac{1}{2} + \left[\frac{1}{\eta^2}\{1 + (\eta - 1)e^{\eta}\}\right]$$

where

$$\eta = \frac{q}{kT}\sqrt{\frac{qE_{\text{ext}}}{\pi\varepsilon_{\text{eff}}}}$$

This model was modified by Jonscher [33], under the assumption that the emission opposite to the electric field should be ignored

$$\aleph_{3D} = \left[\frac{1}{\eta^2}\{1 + (\eta - 1)e^{\eta}\}\right]$$

In these physical models, as the temperature of the silicon increases, the generation rate of carriers will increase. Hence, as silicon self-heating increases, Frenkel–Poole generation will lead to an increase in the generation of carriers and influence the generation rate of the free carriers.

3.2.2.5 Band-to-band tunneling leakage phenomenon

Band-to-band tunneling is important in metallurgical junctions at high electric fields. ESD events can lead to motion of the dopants modifying the nature of the leakage mechanism during ESD events. The issue of tunneling is a greater concern in advanced technologies with the scaling of physical dimensions. Tunneling currents can increase as dielectrics are damaged from ESD events due to changes in the trap state density and establishment of paths of conduction.

Analysis of electron emission in electric fields by quantum mechanics was first performed by R.H. Fowler and L. Nordheim [34]. The Fowler–Nordheim development solved the problem of the transmission through a step barrier in the presence of an applied electric field by Schroedinger's equation. Assuming a step potential at the

intercept, the electric field distorts the step potential to a rectangular barrier. Assuming a step barrier height V_0, a particle of energy E and an applied electric field F, the time-dependent Schroedinger equation is

$$\frac{d^2\Psi}{dx^2} + \kappa^2(E - V_0 + Fx)\Psi = 0 \qquad x > 0$$

$$\frac{d^2\Psi}{dx^2} + \kappa^2(E)\Psi = 0 \qquad x < 0$$

By transformation of variables, the equation can be reduced to a second-order ordinary differential equation with a variable coefficient. Let us define a variable

$$y = \left(-\frac{V_0 - E}{F} + x\right)(\kappa^2 F)^{1/3}$$

then

$$\frac{d^2\Psi}{dy^2} + y\Psi = 0$$

The solution is a Bessel function of order 1/3 where

$$\Psi = \sqrt{y}\{J_{1/3}(2/3y^{3/2})\} \text{ and } \Psi = \sqrt{y}\{J_{-1/3}(2/3y^{3/2})\}$$

Far from the step, the solution must reduce to a freely propagating wave packet, hence this is best expressed as Hankel's second function

$$\Psi = \sqrt{y}\{H^{(2)}{}_{1/3}(2/3y^{3/2})\}$$

From this development, the transmission probability through a triangular barrier with particle energy E is expressed as

$$T(E) \propto \exp\left[-\frac{4}{3}\frac{\sqrt{2m}(E)^{3/2}}{q\hbar F}\right]$$

The Fowler-Nordheim development does not assume a semiconductor material.

Zener addressed the problem for a semiconductor in 1933 [35]. Zener first proposed that a band-to-band tunneling occurs in a semiconductor. The derivation assumes a tunneling mechanism is a one-electron process in a uniform one-dimensional electric field. The transmission probability is determined by the probability that the carrier exists in the conduction band normalized to the probability that it is in the valence band. This is associated with a rate at which the electron impinges on the band edges. Using Bloch theory, Zener showed the rate at which the carrier impinged on the band edge. The velocity of the carrier in k-space is $2\pi eF/h$ where F is the local external field. The width of the Brillouin zone is $2\pi/h$, thus the frequency that the carrier impinges is eFa/h. Zener realized that Bloch wave function analysis assumes real values of wave vectors for

carriers in the band. Generalizing to complex wave vectors allows for an exponentially decaying wave function in the forbidden zone. The generalization of the Bloch wave function was

$$\Psi_k(x) = U_k(x) \exp\left\{ \int iK(x)\, \mathrm{d}x \right\}$$

The transmission probability is then the ratio of the square of the wavefunction at the valence band normalized to the square of the wavefunction at the conduction band. Then

$$\left| \frac{\Psi_v}{\Psi_c} \right|^2 = \exp\left\{ -2 \int_v^c \eta(x)\, \mathrm{d}x \right\}$$

where the integral is solved from the valence band to the conduction band over the imaginary part of the value of $K(x)$. Hence the total tunneling probability is the product of the carrier rate of impinging on the conduction band and the probability of transmission.

$$\frac{eFa}{h} \exp\left\{ -2 \int_v^c \eta(x)\, \mathrm{d}x \right\}$$

To solve for the Zener tunneling rate, the wave function must be solved using Schroedinger's equation in the lattice. Zener solved this problem using a one-dimensional lattice potential and a spatially varying electric field between the two edges. Solving the equation, the value of $K(x)$ is

$$K(x) \simeq \frac{\pi}{a} \left[1 - i \frac{8ma^2}{h^2} \{ V_0^2 - (eFx)^2 \}^{1/2} \right]$$

then the transmission probability is solved in the integral and the transmission rate is

$$\frac{eFa}{h} \exp\left\{ -2 \int_{-V_0/eF}^{V_0/eF} \frac{8ma^2}{h^2} \{ V_0{}^2 - (eFx)^2 \}^{1/2} \mathrm{d}x \right\}$$

By symmetry of the integral,and letting $E_g = 2V_0$

$$T(F) \simeq \frac{eFa}{h} \exp\left\{ -\frac{\pi^2 ma}{h^2} \frac{E_g^2}{eF} \right\}$$

This development was extended by Mcafee, Ryder, Schockley, and Sparks by addressing the effective mass in a semiconductor [36]. Kane, Price and Radcliffe, Keldysh addressed extensions of this model to indirect bandgap silicon structures. In these developments, the phonon emission, and energy losses in the system were compared with the direct gap analysis [37–40]. In the Kane–Price–Keldysh (KPK) developments, the tunneling probability is lower in the indirect bandgap media compared with the

direct bandgap material in the prefactor and in the exponential. In the indirect bandgap materials, some of the tunneling energy is transmitted into phonon energy.

In these developments, the tunneling current is a strong function of electric field and a weak function of the temperature. Self-heating has only a small effect on the tunneling current. From band-to-band tunneling measurements from MOSFET gate-induced drain leakage (GIDL), the current does increase with temperature in the tunneling regime. Hence self-heating will lead to some increase in the tunneling current, but the thermally generated current will increase at a much more significant rate with temperature. As the self-heating from ESD occurs, the thermally generated mechanisms will dominate over the tunneling mechanisms at higher voltages.

3.2.2.6 Avalanche multiplication with electron and hole ionization

Impact ionization leads to the generation of carriers, leading to a cascade effect in a region with an electric field applied [41–43]. Avalanche multiplication occurs for both the negative and positive species when the energy level exceeds the impact ionization threshold for both species.

Let there be an incident current I_{p0} and total current I which is the sum of the positive and negative specie (e.g. electron and hole in semiconductors). At the position $x = W$, the hole current is $M_p I_p$. The differential generation in an increment is associated with the product of the impact ionization coefficient and current at the position in space at that location in space.

$$\mathrm{d}I_p(x) = \{\alpha_p I_p + \alpha_n I_n\}\,\mathrm{d}x$$

This can be expressed as a function of the total current

$$\frac{\mathrm{d}I_p(x)}{\mathrm{d}x} = \alpha_n I + (\alpha_p - \alpha_n) I_p$$

or equivalently

$$\frac{\mathrm{d}I_p(x)}{\mathrm{d}x} - (\alpha_p - \alpha_n) I_p(x) = \alpha_n I$$

In this form, the equation is a first-order ordinary differential equation with a variable coefficient. The integration factor for the integral can be expressed as

$$\mu(x) = \exp\left[-\int_0^x (\alpha_p - \alpha_n)\,\mathrm{d}t\right]$$

with a general solution of

$$I_p(x) = \frac{1}{\mu(x)}\left[\int_0^x \mu(s)(\alpha_n I)\,\mathrm{d}s + C\right]$$

where C is a constant. Then we obtain,

$$I_p(x) = \exp\left[\int_0^x (\alpha_p - \alpha_n)\,\mathrm{d}t\right]\left[\int_0^x \exp\left[-\int_0^s (\alpha_p - \alpha_n)\,\mathrm{d}t\right](\alpha_n I)\,\mathrm{d}s + C\right]$$

This can be expressed as a function of the total current as

$$I_p(x) = I\exp\left[\int_0^x (\alpha_p - \alpha_n)\,\mathrm{d}t\right]\left[\int_0^x \alpha_n \exp\left[-\int_0^s (\alpha_p - \alpha_n)\,\mathrm{d}t\right]\mathrm{d}s + K\right]$$

Letting the hole current at $x = 0$ be I_{p0}, we can solve for the constant where $I_{p0} = M_p I = K$ which is the hole multiplication times the total current at the electrode at $x = 0$. Then $K = 1/M_p$. The general solution is then,

$$I_p(x) = I\exp\left[\int_0^x (\alpha_p - \alpha_n)\,\mathrm{d}t\right]\left[\int_0^x \alpha_n \exp\left[-\int_0^s (\alpha_p - \alpha_n)\,\mathrm{d}t\right]\mathrm{d}s + \frac{1}{Mp}\right]$$

From this form, we can solve for the avalanche multiplication term M_p from the boundary condition at the position $x = W$, where $I_p(W) = M_p I_{p0}$.

$$\frac{1}{M_p} = \exp\left[-\int_0^{x=W} (\alpha_p - \alpha_n)\,\mathrm{d}x\right] - \left[\int_0^{x=W} \alpha_n \exp\left[-\int_0^{x'=x} (\alpha_p - \alpha_n)\,\mathrm{d}x'\right]\mathrm{d}x\right]$$

To integrate the integral expression, changing the form

$$\left[\int_0^{x=W} (\alpha_n - \alpha_p)\exp\left[-\int_0^{x'=x} (\alpha_p - \alpha_n)\,\mathrm{d}x'\right]\mathrm{d}x\right] = \exp\left[-\int_0^W (\alpha_p - \alpha_n)\,\mathrm{d}x\right] - 1$$

Hence by adding the impact ionization term for holes, we obtain an expression equal to the first term of the RHS. Adding this to the equation, and then subtracting out the common integral terms, the multiplication term can be placed in the form as follows

$$1 - \frac{1}{M_p} = \left[\int_0^{x=W} \alpha_p \exp\left[-\int_0^{x'=x} (\alpha_p - \alpha_n)\,\mathrm{d}x'\right]\mathrm{d}x\right]$$

Let us establish a relationship between the hole and electron impact ionization terms,

$$\alpha_p = \gamma\alpha_n$$

$$1 - \frac{1}{M_p} = \frac{\gamma}{1-\gamma}\left\{\exp\left[-\left(1 - \frac{1}{\gamma}\right)\int_0^W (\alpha_p)\,\mathrm{d}x\right] - 1\right\}$$

and

$$1 - \frac{1}{M_n} = \frac{1}{\gamma - 1}\left\{\exp\left[\int_0^W (\gamma - 1)(\alpha_n)\,\mathrm{d}x\right] - 1\right\}$$

From this formulation, we obtain a modified Townsend criterion of the form

$$\int_0^W \alpha_n\,\mathrm{d}x = \frac{\ln\gamma}{\gamma - 1}$$

$$\int_0^W \alpha_p\,\mathrm{d}x = \frac{\gamma\ln\gamma}{\gamma - 1}$$

3.3 BIPOLAR HIGH-CURRENT DEVICE PHYSICS

3.3.1 Bipolar Transistor Equation

Bipolar transistors are present as both active devices and as parasitic elements in integrated technology. During electrostatic discharge, electrical overstress, and latchup, the parasitic bipolar elements play a key role in the understanding of how circuits respond. Parasitic bipolar elements of both *npn* and *pnp* transistor types exists within circuit elements, between adjacent elements, and between circuits and circuit blocks. Parasitic bipolar transistors are inherent in bulk and silicon-on-insulator (SOI) MOSFET devices in parallel with the surface channel MOSFET conduction. To understand bipolar transistors, and parasitic elements in integrated technology it is important to understand bipolar device operation.

Figure 3.5 is an example of a modern epitaxial base silicon–germanium heterojunction bipolar transistor. Today, these transistors have cutoff frequencies well above the frequency response of ESD events.

Evaluation of bipolar transistor current density from the emitter to the collector for an *npn* transistor can be represented as [23]

$$J = \frac{qn_i^2\left[\exp\left(\dfrac{qV_{\mathrm{BC}}}{kT}\right) - \exp\left(\dfrac{aV_{\mathrm{BE}}}{kT}\right)\right]}{\displaystyle\int_0^{x_B}\frac{p(x)}{D_n(x)}dx}$$

where we can express the current in the form,

$$I = I_{\mathrm{s}}\left[\exp\left(\frac{qV_{\mathrm{BC}}}{kT}\right) - \exp\left(\frac{qV_{\mathrm{BE}}}{kT}\right)\right]$$

$$I_{\mathrm{s}} = \frac{q^2A^2n_i^2\langle D_n\rangle}{Q_{\mathrm{B}}}$$

$$\langle D_n\rangle = \frac{qA\displaystyle\int_0^{x_B} p(x)\,\mathrm{d}x}{\displaystyle\int_0^{x_B}\frac{p(x)}{D_n}\,\mathrm{d}x}$$

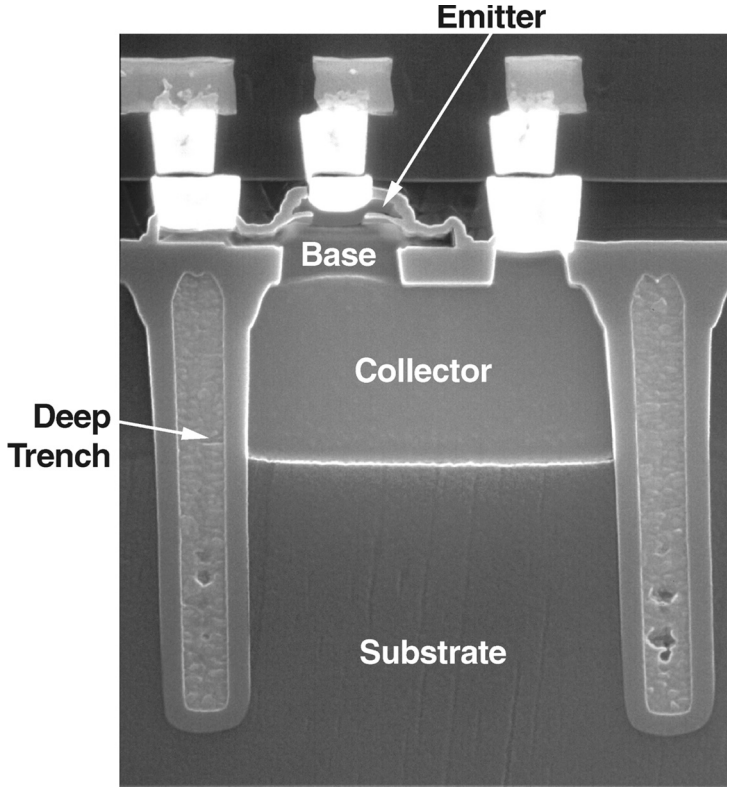

Figure 3.5 Cross-section of a silicon–germanium epitaxial bipolar transistor

and

$$Q_{\mathrm{B}} = qA \int_0^{x_B} p(x)\,\mathrm{d}x$$

The base current can be expressed in a similar form where current is a function of the excess charge in the base divided by the recombination time,

$$Q'_{\mathrm{B}} = qA \int_0^{x_{\mathrm{B}}} \left[n(x) - n_0 \right] \mathrm{d}x$$

$$I_{\mathrm{rB}} = \frac{Q'_{\mathrm{B}}}{\tau_n}$$

3.3.2 Kirk Effect

During an ESD event, high current will flow in a bipolar transistor, in bipolar transistor peripheral circuits, or ESD power clamps. At these high currents, the frequency response of the transistor is impacted by the current flowing in the base region. As the

current density increases, the space-charge region in the base–collector junction is pushed outward. This effect is known as the Kirk effect [44]. In analysis of the electrostatics in a junction it is assumed that the carriers are swept out of the depletion region and do not modulate the electric field. At high currents, the ESD current flowing through the junction modifies the space-charge region, leading to a space-charge region modulation.

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{1}{\varepsilon}\left[qN(x) - \frac{J_\mathrm{c}}{v(x)}\right]$$

where J_c is the current density from the ESD current flowing through the transistor element. Assuming that the collector–base voltage is a constant, then the collector–base voltage and the built-in potential will equal the voltage drop across the space-charge layer. The voltage drop across the depletion region is the integral of the electric field over the junction from the collector to base region. This can be expressed as

$$V_\mathrm{CB} = \frac{1}{\varepsilon}\int_{x_\mathrm{b}}^{x_\mathrm{c}} x\left[qN(x) - \frac{J_\mathrm{c}}{v(x)}\right]\mathrm{d}x - \phi_i$$

This can be simplified by assuming the doping concentration is a constant in the epitaxial region,

$$V_\mathrm{CB} = \frac{1}{2\varepsilon}\left[qN_\mathrm{epi} + \frac{J_\mathrm{c}}{v_\mathrm{sat}}\right](x_\mathrm{c} - x_\mathrm{b})^2 - \phi_i$$

Defining a saturation current

$$J_\mathrm{sat} = qN_\mathrm{epi}v_\mathrm{sat}$$

then

$$V_\mathrm{CB} = \frac{qN_\mathrm{epi}}{2\varepsilon}\left[1 + \frac{J_\mathrm{c}}{J_\mathrm{sat}}\right](x_\mathrm{c} - x_\mathrm{b})^2 - \phi_i$$

From this the variation of the base width as a function of the ESD current can be solved as

$$W_\mathrm{CB}(J_\mathrm{ESD}) = \frac{W_\mathrm{CB0}}{\left(1 + \dfrac{J_\mathrm{ESD}}{J_\mathrm{sat}}\right)^{1/2}}$$

where

$$W_\mathrm{CB0} = \left[\frac{2\varepsilon(V_\mathrm{CB} + \phi_i)}{qN_\mathrm{epi}}\right]^{1/2}$$

The ESD collector-to-emitter current will flow through the collector region. As the ESD current increases, the collector–base depletion width will decrease and the base width will increase, leading to a lower response to the ESD event, if the ESD event is

slower than the charge relaxation time. When the ESD event achieves the peak current, the transistor response will be least responsive. Hence we can express a time-dependent depletion width relationship during an ESD event as

$$W_{\mathrm{CB}}(t) = \frac{W_{\mathrm{CB0}}}{\left(1 + \dfrac{J_{\mathrm{ESD}}(t)}{J_{\mathrm{sat}}}\right)^{1/2}}$$

where the depletion width is a function of the ESD pulse width time response.

3.3.3 Johnson Limit

A fundamental relationship exists between the frequency response of the transistor and the maximum power applied across a transistor. This is known as the Johnson limit [45]. Early, Pritchard and Johnson all addressed the physical limitation of transistor frequency response and power [45–50].

This is important to ESD phenomena for two reasons. The response time of the transistor relative to the ESD pulse influences the time constant hierarchy ordering. The effect in which the element responds to the ESD pulse will influence the operation and response to ESD events. Second, the speed of the transistor is related to the breakdown voltages.

Johnson addressed the fundamental relationship of how much voltage can you place across a gap and accelerate an electron. Given a gap of width W, one can apply a maximum voltage across the region

$$V_{\mathrm{max}} = E_{\mathrm{max}} W$$

To traverse a gap of width W, this is equal to the product of the drift velocity of the electron in the electric field and the transit time τ_{TR} to traverse the gap W. The maximum electric field can be understood as the critical field E_{c} associated with electron velocity saturation v_{sat}. Substituting

$$V_{\mathrm{max}} = E_{\mathrm{max}}(v_{\mathrm{sat}})\tau_{\mathrm{TR}}$$

The transit time can be equated to the fastest frequency at which we can transport an electron. This transit time, in the case of a bipolar transistor, is the time required to traverse the base region, which can be regarded as related to the unity current gain cutoff frequency f_{T} (ignoring the emitter and collector transit time), hence simplistically

$$\tau_{\mathrm{TR}} = \frac{1}{2\pi f_{\mathrm{T}}}$$

The maximum electric field that can be placed across the medium is the breakdown field V_{BR}. Substituting in the variables, we obtain

$$V_{\mathrm{BR}} f_{\mathrm{T}} = \frac{E_{\mathrm{m}} v_{\mathrm{sat}}}{2\pi}$$

In this form, it states that there is an inverse relationship between the maximum power and frequency response. The formulation states that the product of the maximum velocity an electron can traverse a medium and the maximum electric field across that region is a constant. It also states that there is an inverse relationship between the transistor speed and the allowed breakdown voltage (Figure 3.6).

Based on the Johnson Limit condition, as a technology scales to provide a higher unit current gain cutoff frequency f_T, the BV_{CEO} of the transistor decreases. Hence from the Johnson limit equation,

$$V^*_m f^*_T = V_m f_T = \frac{E_m v_s}{2\pi}$$

where $V^*_m f^*_T$ is associated with a first transistor and $V_m f_T$ is associated with a second transistor. The ratio of breakdown voltages can be determined as

$$\frac{V^*_m}{V_m} = \frac{f_T}{f^*_T}$$

The Johnson limit can be represented in different forms. In the power formulation, we can relate the maximum power that can be transmitted across a gap W. The Johnson limit in its power formulation is given as

$$(P_m X_c)^{1/2} f_T = \frac{E_m \upsilon_s}{2\pi}$$

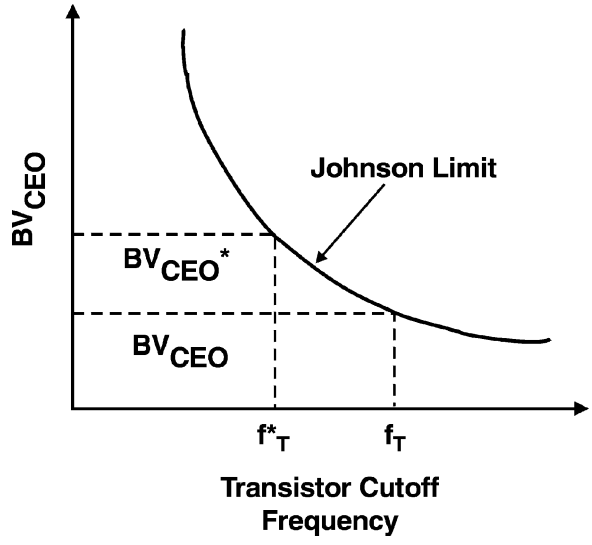

Figure 3.6 Plot of the breakdown voltage (BV_{CEO})–unit current gain cutoff frequency (f_T) relationship for bipolar transistors (Johnson limit relationship)

where P_m is the maximum power and X_c is the reactance

$$X_c = \frac{1}{2\pi f_T C_{bc}}$$

In this form, it states that there is an inverse relationship between the maximum power and reactance. From the power formulation, there exists a maximum power that can be transmitted across the bipolar transistor base. From an ESD perspective, there is also a maximum power to failure of the physical region for a given pulse width. From this derivation, we can compare the power to failure and the maximum power according to the Johnson limit equation. This can be put in the form of a dimensionless group, which represents the thermal property response to the electrical response.

From the ratio of the power to failure condition of the Wunsch–Bell curve and the maximum power – f_T relationship, a dimensionless group can be established explaining the relationship between thermal conduction, thermal capacity, failure temperature, pulse width, saturation velocity, maximum electric field condition and the unity current gain cutoff frequency. Defining a dimensionless group V_0

$$V_0 = \frac{\text{Power to failure}}{\text{Johnson limit maximum power}} = \frac{P_f(\tau)}{P_m}$$

The Wunsch–Bell model and the Johnson model are both one-dimensional. In the thermal diffusion time regime, we can express dimensionless group V_0 as

$$V_0 = \frac{P_f}{P_m} = \frac{A\sqrt{\dfrac{\pi K C_p \rho (T - T_0)}{\tau}}}{\dfrac{\left(\dfrac{E_m^2 v_s^2}{(2\pi)^2}\right)}{X_c f_T^2}}$$

3.4 SILICON-CONTROLLED RECTIFIERS

Silicon-controlled rectifier (SCR) structures (Figure 3.7) are important semiconductor devices for electrostatic discharge, latchup and power electronics [51–57]. Silicon-controlled rectifiers have many applications in power electronics as high current switches. Silicon controlled rectifiers are also used as electrostatic discharge protection networks on input nodes, and on power rails as a low-impedance shunt device. Additionally, the SCR is also an inherent parasitic structure in semiconductor technologies, which can lead to latchup. These structures, whether intentionally constructed or as an unintentional parasitic device, is important to quantify and understand for ESD protection, and latchup analysis.

The SCR device plays a role as a high-current switch as a result of its S-type I–V characteristic allows it to transition from a low-current/high-voltage state to a high-current/low-voltage state. It is this feature which serves it well as an electrostatic

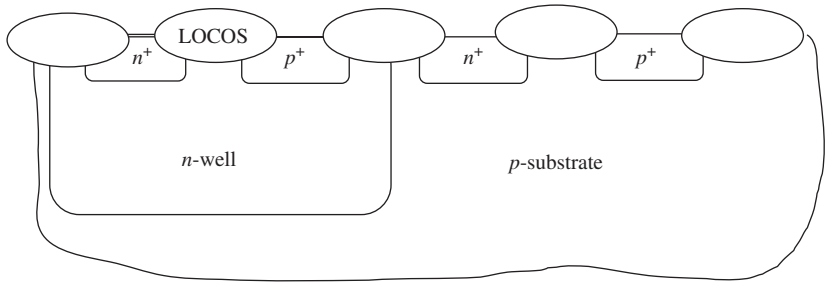

Figure 3.7 Silicon-controlled rectifier structure

discharge network. It is 'off' in normal operation, and can be triggered 'on' into a high-current state. In this state, it establishes a high current at a low voltage, allowing a low-impedance shunt. The silicon-controlled rectifier, also known as the Shockley diode, is a four-region device of alternating *p*- and *n*-doped regions with three physical *p*–*n* metallurgical junctions.

Conceptually the four-region *pnpn* can be understood as a cross-coupled *pnp* and *npn* bipolar junction transistor (BJT) device, where the base of the *pnp* BJT device is the collector of the *npn* BJT device, and the base of the *npn* is the collector of the *pnp* BJT device. This *pnp*–*npn* BJT coupling establishes regenerative feedback leading to the S-type *I*–*V* characteristic and the cause of the electrical instability which is observed as a negative resistance state ($\mathrm{d}I/\mathrm{d}V < 0$). It is this feature that has made the Shockley diode a valuable device for power electronics, and ESD protection as well as its danger as an enabler of CMOS latchup (Figure 3.8).

Application of a positive bias on the emitter of the *pnp* element and a ground potential on the emitter of the *npn* element establishes a voltage across the *pnpn*. The positive voltage serves as to provide a forward biasing of the emitter–base junctions of the *pnp* and *npn* transistors. The base–collector junction of the *pnp* (which is also the base–collector junction of the *npn*) is in a reverse-biased state. This prevents current flow from the anode of the *pnpn* to the cathode. As the voltage is increased, the voltage across the base–collector junction increases. This mode of operation is called the forward blocking state. In order for current to flow efficiently from the *pnpn* anode to the cathode, the base–collector junction must allow current to flow.

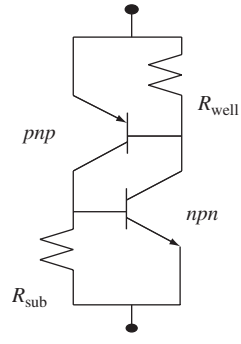

Figure 3.8 Silicon-controlled rectifier circuit schematic with shunt resistors

For current continuity at the cross-coupled nodes, the collector current of the *pnp* transistor must equal the base current of the *npn* transistor, as well as the collector current of the *npn* transistor must equal the base current of the *pnp* transistor. Mathematically, the coupling is established through solving Kirchoff's current law at the base–collector nodes. In this form, the standard equations of bipolar transistors can be used to quantify the interaction and current in the *pnpn* structure. Hence the two nodal equations can be quantified and expressed as

$$I_{cp} = I_{bn}$$
$$I_{cn} = I_{bp}$$

where I_{cp} and I_{cn} are the collector current of the *pnp* and *npn* BJT, respectively, and likewise, I_{bn} and I_{bp} are the base current of the *npn* and *pnp* BJT, respectively.

The total current through the *pnpn* structure is equal to the emitter current of the *pnp* or the emitter current of the *npn* BJT transistors. From Kirchoff's current law in the transistor, the emitter current must equal the base and collector current.

$$I = I_{ep} = I_{cp} + I_{bp}$$

From the coupling relationships, we can express this as

$$I = I_{ep} = I_{cp} + I_{cn} = I_{bp} + I_{bn} = I_{cn} + I_{bn} = I_{en} = I$$

Solving for the current as a function of the two collector relationships, we can relate the collector current as a function of the emitter current

$$I_{cp} = \alpha I_{ep} + I_{cp0}$$
$$I_{cn} = \alpha I_{en} + I_{cn0}$$

where the collector current is equal to the transport factor times the emitter current and the base–collector leakage. Solving for the current through the *pnpn* structure, we obtain

$$I = \frac{I_{cp0} + I_{cn0}}{1 - (\alpha_n + \alpha_p)}$$

This can be expressed as a function of the bipolar transistor current gain, substituting in for the collector–emitter transport

$$I = \frac{I_{cp0} + I_{cn0}}{1 - \left(\dfrac{\beta_n}{\beta_n + 1} + \dfrac{\beta_p}{\beta_p + 1}\right)}$$

The current magnitude in the SCR device can be large, given that the numerator is large, or if the denominator approaches zero. This physically occurs if the leakage current in the base–collector junction is large due to impact ionization, punch-through,

photons, or an external source of current. This is also a function of the bipolar current gain values. In a BJT device, avalanche breakdown occurs at,

$$M\alpha = 1$$

and a generalization to the *pnpn* structure, it is evident from the previous equation that avalanche will occur when

$$M_p\alpha_p + M_n\alpha_n = 1$$

3.4.1 Regenerative Feedback

The condition for triggering of the silicon-controlled rectifier (SCR) is a function of the currents through the feedback elements. The intrinsic resistances in the collector, base and emitter structure can influence the regenerative feedback condition. The evaluation of the base resistance plays a critical role in modulation of the regenerative feedback. The base resistances, whether intentional or unintentional, play a significant role in holding voltage in a SCR structure. For avoidance of unintentional SCR action in CMOS technology, the well and substrate resistances serve as the base resistances. These resistances can be modified by both semiconductor process and design.

It has been shown that the total SCR current equals the emitter current. The emitter current is equal to the sum of the collector currents, or the sum of the base currents.

$$I = I_{cp} + I_{cn}$$

and

$$I_{cp} = \alpha I_{ep} + I_{cp0}$$
$$I_{cn} = \alpha I_{en} + I_{cn0}$$

The base resistance terms of the *pnp* and the *npn* transistor are outside of the collector–base current loop in the regenerative feedback, but they do play a role in the relationship between the emitter current and the current through the SCR structure.

$$I = I_{ep} + I_{w}$$
$$I = I_{en} + I_{sx}$$

Then

$$I = I_{cp} + I_{cn} = \alpha_p I_{ep} + I_{cp0} + \alpha_n I_{en} + I_{cn0}$$
$$I = \alpha_p(I_{ep} + I_{w}) - \alpha_p(I_{w}) + I_{cp0} + \alpha_n(I_{en} + I_{sx}) - \alpha_n(I_{sx}) + I_{cn0}$$
$$I = (\alpha_p + \alpha_n)I - \alpha_p(I_{w}) - \alpha_n(I_{sx}) + (I_{cn0} + I_{cp0})$$
$$I = \frac{(I_{cn0} + I_{cp0}) - \alpha_p(I_{w}) - \alpha_n(I_{sx})}{1 - (\alpha_p + \alpha_n)}$$

From the equation for the total current relationship

$$(\alpha_p + \alpha_n)I - \alpha_p(I_w) - \alpha_n(I_{sx}) + (I_{cn0} + I_{cp0}) - I = 0$$

dividing by the current

$$(\alpha_p + \alpha_n) - \alpha_p \frac{(I_w)}{I} - \alpha_n \frac{(I_{sx})}{I} + \frac{(I_{cn0} + I_{cp0})}{I} - 1 = 0$$

Assuming that the leakage current is significantly less than the current, we can express the condition for latchup as

$$(\alpha_p + \alpha_n) = 1 + \alpha_p \frac{(I_w)}{I} + \alpha_n \frac{(I_{sx})}{I}$$

This can be expressed as a function of the forward bipolar current gains

$$\beta_p\beta_n = 1 + \frac{(I_w)}{I}\beta_p(\beta_n + 1) + \frac{(I_{sx})}{I}(\beta_p + 1)\beta_n$$

To determine the bipolar current gain required to satisfy this condition, knowing one of the bipolar current gains, we can solve for the current gain of the second transistor as a function of the first transistor. Factoring the above equation

$$\beta_p\left[\beta_n - \frac{(I_w)}{I}(\beta_n + 1) - \frac{(I_{sx})}{I}\beta_n\right] = 1 + \frac{(I_{sx})}{I}\beta_n$$

Factoring and solving for the bipolar current gain of the *pnp* transistor needed as a function of the bipolar current gain of the *npn* transistor, well resistor current, substrate resistance current and the current through the SCR [55–57],

$$\beta_p \geqslant \frac{\dfrac{(I)}{\beta_n} + I_{sx}}{\left[I - I_w\dfrac{(\beta_n + 1)}{\beta_n} - I_{sx}\right]}$$

To obtain the criteria as a function of the product of the current gains, we can multiply the expression by the *npn* current gain,

$$\beta_n\beta_p \geqslant \frac{I + I_{sx}\beta_n}{\left[I - I_w\dfrac{(\beta_n + 1)}{\beta_n} - I_{sx}\right]}$$

or equivalently due to symmetry

$$\beta_n\beta_p \geqslant \frac{I + I_w\beta_p}{\left[I - I_{sx}\dfrac{(\beta_p + 1)}{\beta_p} - I_w\right]}$$

The well and substrate currents can be solved for as a function of the base–emitter voltage divided by the well and substrate resistance, respectively.

$$I_{\mathrm{w}} = \frac{(V_{\mathrm{BE}})_{pnp}}{R_{\mathrm{w}}} = \frac{V_0}{R_{\mathrm{w}}} \ln\left[\frac{I - I_{\mathrm{w}}}{(I_0)_p}\right]$$

$$I_{\mathrm{sx}} = \frac{(V_{\mathrm{BE}})_{npn}}{R_{\mathrm{sx}}} = \frac{V_0}{R_{\mathrm{sx}}} \ln\left[\frac{I - I_{\mathrm{sx}}}{(I_0)_n}\right]$$

3.4.1.1 Regenerative feedback with emitter resistance

The intrinsic resistances in the emitter structure can influence the regenerative feedback condition. Emitter resistances influence the electrical and thermal stability of a network (Figure 3.9). Emitter resistance also influences the latchup stability of a network and influences the latchup criteria. The condition for triggering of the silicon controlled rectifier (SCR) is a function of the currents through the feedback elements.

To determine the bipolar current gain required to satisfy this condition, knowing one of the bipolar current gains, we can solve for the current gain of the second transistor as a function of the first transistor, factoring and solving for the bipolar current gain of the *pnp* transistor needed as a function of the bipolar current gain of the *npn* transistor, well resistor current, substrate resistance current and the current through the SCR. With the presence of the emitter resistances, the substrate and well resistances are modified

$$\beta_p \geqslant \frac{\dfrac{(I)}{\beta_n} + I'_{\mathrm{sx}}}{\left[I - I'_{\mathrm{w}} \dfrac{(\beta_n + 1)}{\beta_n} - I'_{\mathrm{sx}}\right]}$$

$$\beta_n \beta_p \geqslant \frac{I + I'_{\mathrm{sx}} \beta_n}{\left[I - I'_{\mathrm{w}} \dfrac{(\beta_n + 1)}{\beta_n} - I'_{\mathrm{sx}}\right]}$$

or equivalently due to symmetry

$$\beta_n \beta_p \geqslant \frac{I + I'_{\mathrm{w}} \beta_p}{\left[I - I'_{\mathrm{sx}} \dfrac{(\beta_p + 1)}{\beta_p} - I'_{\mathrm{w}}\right]}$$

The emitter resistance in the *pnp* transistor influences the voltage drops across the *n*-well resistor. Using Kirchoff's voltage law, the sum of the voltage drops in the current loop across the two resistor elements (well and emitter) and emitter–base voltage equals zero. The current flowing through the emitter resistor is the SCR current minus the current through the *n*-well resistor, hence the well resistance current is

$$I'_{\mathrm{w}} = \frac{(V_{\mathrm{BE}})_{pnp} + I r_{\mathrm{e}p}}{R_{\mathrm{w}} + r_{\mathrm{e}p}}$$

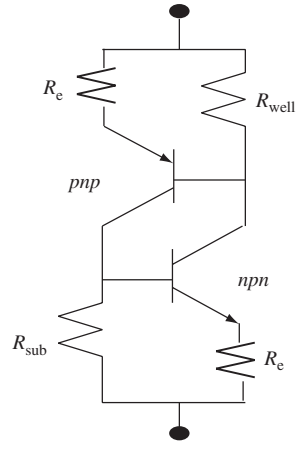

Figure 3.9 Silicon-controlled rectifier circuit schematic with emitter resistance

Equivalently, the substrate resistance current is

$$I'_{sx} = \frac{(V_{BE})_{npn} + Ir_{en}}{R_{sx} + r_{en}}$$

For the case of the emitter resistance for the *npn* (ignoring the *pnp* emitter resistance) [55,56],

$$\beta_n\beta_p \geqslant \frac{I + \left(\dfrac{(V_{BE})_{npn} + Ir_{en}}{R_{sx} + r_{en}}\right)\beta_n}{\left[I - I'_w\dfrac{(\beta_n + 1)}{\beta_n} - \left(\dfrac{(V_{BE})_{npn} + Ir_{en}}{R_{sx} + r_{en}}\right)\right]}$$

For the case of the emitter resistance for the *pnp* (ignoring the *npn* emitter resistance)

$$\beta_n\beta_p \geqslant \frac{I + \left(\dfrac{(V_{BE})_{pnp} + Ir_{ep}}{R_w + r_{ep}}\right)\beta_p}{\left[I - I'_{sx}\dfrac{(\beta_p + 1)}{\beta_p} - \left(\dfrac{(V_{BE})_{pnp} + Ir_{ep}}{R_w + r_{ep}}\right)\right]}$$

or addressing both the emitter resistance of the *pnp* and the *npn* transistor [56]

$$\beta_n\beta_p \geqslant \frac{I + \left(\dfrac{(V_{BE})_{npn} + Ir_{en}}{R_{sx} + r_{en}}\right)\beta_n}{\left[I - \left(\dfrac{(V_{BE})_{pnp} + Ir_{ep}}{R_w + r_{ep}}\right)\dfrac{(\beta_n + 1)}{\beta_n} - \left(\dfrac{(V_{BE})_{npn} + Ir_{en}}{R_{sx} + r_{en}}\right)\right]}$$

3.4.1.2 Holding current

The condition for holding current in a silicon controlled rectifier can be obtained from the triggering condition criteria. From the derivation of the trigger condition inequality, it can be shown that when the expression is equal, the current is the holding current. Let

$$\beta_p\beta_n = 1 + \frac{(I_w)}{I}\beta_p(\beta_n + 1) + \frac{(I_{sx})}{I}(\beta_p + 1)\beta_n$$

Multiplying both sides by the current I, and factoring out the terms,

$$I_H = \frac{\beta_p(\beta_n + 1)I_w + (\beta_p + 1)\beta_n I_{sx}}{\beta_p\beta_n - 1}$$

where the well and substrate currents can be solved for as a function of the base–emitter voltage divided by the well and substrate resistance, respectively.

$$I_w = \frac{(V_{BE})_{pnp}}{R_w} = \frac{V_0}{R_w}\ln\left[\frac{I - I_w}{(I_0)_p}\right]$$

$$I_{sx} = \frac{(V_{BE})_{npn}}{R_{sx}} = \frac{V_0}{R_{sx}}\ln\left[\frac{I - I_{sx}}{(I_0)_n}\right]$$

In the case where the SCR *npn* gain is significantly larger than unity, and the SCR *pnp* gain is small [55–57],

$$I_H|_{\beta n>>1} \cong I_w + \left(\frac{\beta_p + 1}{\beta_p}\right)I_{sx}$$

and the complementary case

$$I_H|_{\beta_p>>1} \cong I_{sx} + \left(\frac{\beta_n + 1}{\beta_n}\right)I_w$$

or in the case that both bipolar current gains are significant,

$$I_H|_{\beta n\beta p>>1} \cong I_{sx} + I_w$$

Hence, the holding current relationship is the sum of the base–emitter voltage of the *pnp* and *npn* transistor divided by the well, and substrate resistance, respectively.

3.4.1.3 Generalized tetrode condition

Evaluation of the operation of a silicon controlled rectifier and its generalized condition (Figure 3.10) for switching can be evaluated using a stability criteria as a function of the effective large-scale injection factors, and the effective transport factors [57]. Let us

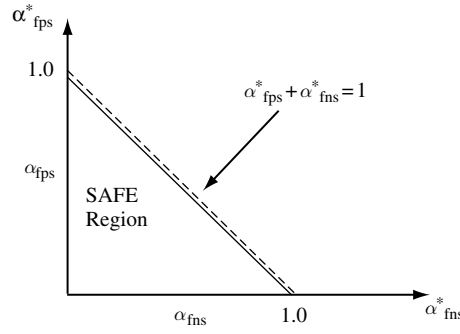

Figure 3.10 Graphical representation of generalized tetrode relationship for stability and instability

define the effective large-scale injection factors for the *pnp* and *npn* as the ratio of current flowing through the emitter compared with the total current flowing through the SCR

$$\gamma_p^* = \frac{I_{ep}}{I}$$
$$\gamma_n^* = \frac{I_{en}}{I}$$

where the I_{ep} and I_{en} are the emitter currents and I is the total SCR current. The total SCR current I is larger due to the current flowing through the bypass well and substrate resistors. The total current flowing through the SCR can be expressed as

$$I = I_n + I_p = \alpha_{fn} I_{en} + \alpha_{fp} I_{ep} + I_{nd} + I_{pd} + I_{sc}$$

where the first term is the collector currents, and the last terms are the total diffusion current at the well junction and the space charge generation current at the well junction. Expressing the current as a function of the effective injection factors

$$I = I_n + I_p = \alpha_{fn}\gamma_n^* I + \alpha_{fp}\gamma_p^* I + (I_{nd} + I_{pd} + I_{sc})$$

Solving for the total current flowing through the SCR, the expression provides the expression

$$I = \frac{(I_{nd} + I_{pd} + I_{sc})}{1 - \left[\alpha_{fn}\gamma_n^* + \alpha_{fp}\gamma_p^*\right]}$$

Defining an effective bipolar transistor gain as

$$\alpha_{fn}^* = \alpha_{fn}\gamma_n^*$$
$$\alpha_{fp}^* = \alpha_{fp}\gamma_p^*$$

Expressing the leakages as a summation of the total diffusion current at the well junction and the generation within the well depletion region

$$I_{\mathrm{s}} = I_{n\mathrm{d}} + I_{p\mathrm{d}} + I_{\mathrm{sc}}$$

we can express the stability relationship as the ratio of the total diffusion current and generation within the well region, divided by the effective transistor gains

$$I = \frac{I\mathrm{s}}{1 - \left[\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p}\right]}$$

From this relationship, it appears that the switching state will occur when the denominator is zero. Troutman pointed out that this is cannot occur, and that the usage of this expression for determining the switching state is inaccurate, and that a generalized tetrode stability relationship must be represented as the derivative of this condition. In this form, the differential stability criteria can be established by taking the derivative of the expression with respect to the SCR current.

$$\frac{\mathrm{d}I}{\mathrm{d}I_{\mathrm{s}}} = \frac{1}{1 - \dfrac{\mathrm{d}}{\mathrm{d}I}\left[I\left(\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p}\right)\right]}$$

Hence, the condition for instability of the SCR occurs when the denominator is negative, or

$$\frac{\mathrm{d}}{\mathrm{d}I}\left[I\left(\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p}\right)\right] > 1$$

and stable when

$$\frac{\mathrm{d}}{\mathrm{d}I}\left[I\left(\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p}\right)\right] < 1$$

The condition for instability can be expanded, and expressed as

$$\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p} + \frac{\mathrm{d}}{\mathrm{d}I}\left[\left(\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p}\right)\right] > 1$$

From this form, it can be observed that if the last term is positive and greater than zero, the sum of the effective gains will always be less than unity. Hence, if we are to define a triangular region in $\alpha^*_{\mathrm{f}n}$-$\alpha^*_{\mathrm{f}p}$ space where

$$\alpha^*_{\mathrm{f}n} + \alpha^*_{\mathrm{f}p} \leqslant 1$$

then this is the region where the blocking state of the SCR is such that the device does not trigger, and for the domain outside of the triangular domain, switching can occur and the SCR can be unstable.

3.5 RESISTORS

Resistor physics is important for understanding the response of circuits and ESD networks. Diffused resistors may consist of n^+ diffusions, p^+ diffusions, n-wells, or polysilicon elements. In many cases, the resistor elements have silicide block masks to prevent a low resistance state of the structure. The diffused resistor profile may consist of a dopant source which can be expressed as a complementary error function

$$C(x,t) = C_s \; erfc\left(\frac{x}{2\sqrt{Dt}}\right) = \frac{2C_s}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{Dt}}}^{\infty} e^{-v^2} \, dv$$

The total dopants can be found by integration over the profile

$$N' = \int_0^{\infty} C(x,t) \, dx = 2\sqrt{\frac{Dt}{\pi}} C_s$$

After the hot process time t, the well can be represented with a doping concentration

$$C(x,t) = \frac{N'}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right)$$

In a nonuniform profile, the conductance of the well resistor can be found by integration over the well region from the top surface ($x = 0$) to the metallurgical junction, $x = X_J$

$$G = \frac{W}{L} \int_0^{X_J} q\mu n(x) dx$$

When the background doping concentration is negligible compared with the well implant dose, we can solve for the conductance of the well from the Gaussian representation

$$G = \frac{N'q}{\sqrt{\pi Dt}} \frac{W}{L} \int_0^{X_J} \mu n \left[\exp\left(-\frac{x^2}{4Dt}\right)\right] dx$$

In the case where the surface is silicided, the conductance will be modified by the penetration of the silicide into the diffusion and the net resistance is the parallel resistance of the two films

$$R = R_{sal} \| R_{film}$$

where

$$R_{sal} = \rho_{sal} \frac{L}{W}$$

$$R_{film} = \frac{1}{\frac{N'q}{\sqrt{\pi Dt}} \frac{W}{L} \int_{X_{sal}}^{X_J} \mu n \left[\exp\left(-\frac{x^2}{4Dt}\right)\right] dx}$$

In an unsalicided region under an isolation region such as shallow trench isolation, as in the case of a buried resistor (BR) element, or an *n*-well the conductance can be expressed as

$$G \simeq \frac{N'q}{\sqrt{\pi Dt}} \frac{W}{L} \int_{X_{\mathrm{STI}}}^{X_W} \mu n \left[\exp\left(-\frac{x^2}{4Dt} \right) \right] \mathrm{d}x$$

In the case of a retrograde well, dopants are implanted at a depth below the surface, and a second well implant or background 'fill' implant is formed. Assuming a single well implant, where the fill implant is significantly less than the high energy MeV implant we can approximate the conductance as this can be expressed as

$$G \simeq \frac{W}{L} \int_{X_{\mathrm{STI}}}^{X_W} qN_0\mu_0 + \frac{N'q}{\sqrt{\pi Dt}} \mu_n \left[\exp\left(-\frac{(x - x_0)^2}{4Dt} \right) \right] \mathrm{d}x$$

In a semiconductor, the resistor current density is a function of the drift current component.

$$J_{\mathrm{drift}} = qnv_{\mathrm{d}}$$

In a resistor structure, electric fields can approach the velocity saturation level of an electron. A form of the velocity–electric field relationship can be expressed as [58]

$$v_{\mathrm{d}}(E) = v_{\mathrm{sat}} \frac{(E/E_{\mathrm{c}})}{[1 + (E/E_{\mathrm{c}})^{\beta}]^{1/\beta}}$$

where E is the electric field, E_{c} is the critical electrical field, and β is a parameter. Substitution into the drift equation, we obtain

$$J(E) = qnv_{\mathrm{sat}} \frac{(E/E_{\mathrm{c}})}{[1 + (E/E_{\mathrm{c}})^{\beta}]^{1/\beta}}$$

At high electric fields, the internal electric field approaches the critical electric field. At the critical electric field, roll-off of the current density occurs, with the saturated current density can be expressed as

$$J_{\mathrm{sat}}(E > E_{\mathrm{c}}) \simeq qnv_{\mathrm{sat}}$$

where the doping concentration is the concentration of the resistor element.
Resistor elements will undergo three regions of operation (Figure 3.11). In the linear regime, the device has a linear resistance characteristic. As the device approaches the critical electric field, the resistor undergoes a saturation region. This is followed by an avalanche breakdown regime when the voltage is such that avalanche multiplication occurs. Hower and Reddi represented this phenomenon as a space-charge-neutral region [59], and a space-charge-limited region.

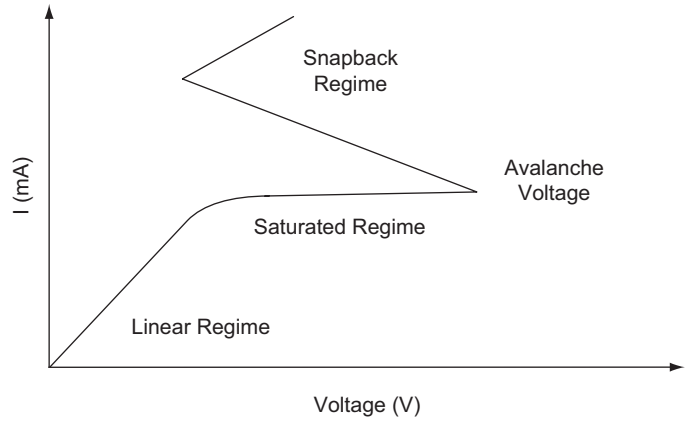

Figure 3.11 Resistor current–voltage characteristic

Making the equation consistent with velocity saturation and avalanche models, let us express the current voltage regimes as [58]

Linear regime

$$I = \frac{V}{R_{\text{lin}}}$$

Saturated regime

$$I = \frac{V}{R_{\text{lin}}\sqrt{1 + \left(\frac{V}{V_{\text{c}}}\right)^2}}$$

Avalanche regime

$$I = \frac{V}{R_{\text{lin}}\sqrt{1 + \left(\frac{V}{V_{\text{c}}}\right)^2}} M$$

$$M = \frac{1}{1 - A\exp\left(-\frac{B}{V_{\text{d}}}\right)}$$

where A, B are physical constants,

$$V_{\text{d}} = \theta V$$

and

$$\theta = 1 - \frac{E_x L}{V}$$

where E_x is the location at which the space charge density transitions into the avalanche multiplication, L is the length of the resistor element, and V is the voltage across the resistor. Puvvuda demonstrates this form of the transition from the linear, to the saturation followed by the avalanche multiplication region.

In this development, self-heating is not shown. Self-heating can occur prior to avalanche breakdown, leading transition into thermal breakdown prior to electrical breakdown. This would be evident in resistor elements which have poor thermal conductivity and low electric fields due to the length.

3.6 MOSFET HIGH-CURRENT DEVICE PHYSICS

3.6.1 Parasitic Bipolar Transistor Equation

MOSFET and MOSFET parameters are important for prediction of ESD sensitivity of both the ESD elements and the circuit networks interfaces. Figure 3.12 shows the key parameters of interest in the MOSFET structure.

Second breakdown is a key subject of interest in the ESD robustness of MOSFET devices. Second breakdown is an electrothermal effect typically leading to thermal runaway, and component failure [60–74]. Parasitic bipolar transistors are inherent in bulk and silicon-on-insulator (SOI) MOSFET devices in parallel with the surface channel MOSFET conduction. In a MOSFET, the current flow from the drain to source. In parallel with the drain and source diffusion, a parasitic bipolar transistor (Figure 3.13) is formed.

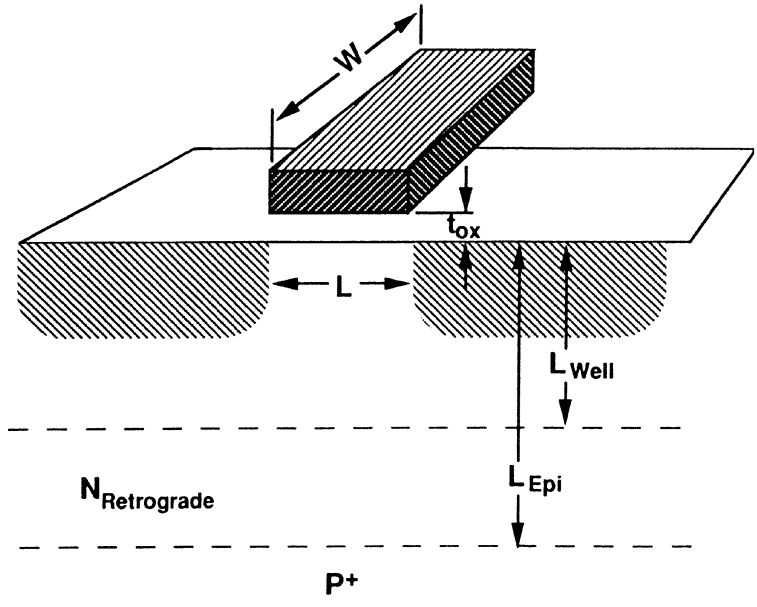

Figure 3.12 MOSFET device

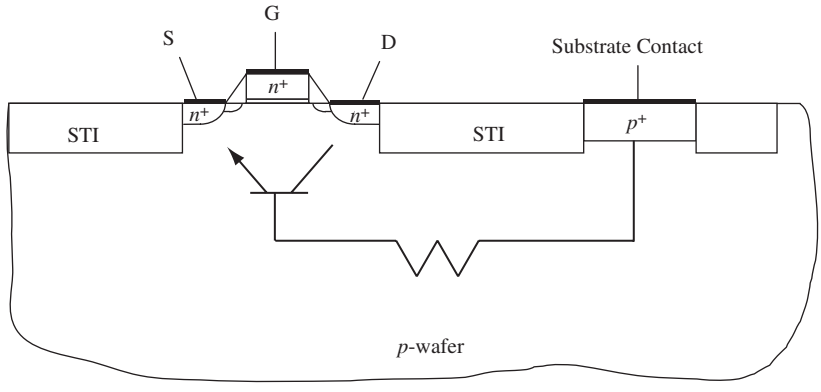

Figure 3.13 MOSFET device with parasitic bipolar transistor

The total drain current can be expressed as a function of the MOSFET drain-to-source current, the parasitic bipolar current, and the thermal generation current [74–81]

$$I = I_{DS} + I_C + I_G$$

The parasitic bipolar transistor can express the current in the form [23]

$$I_C = I_s\left[\exp\left(\frac{qV_{BC}}{kT}\right) - \exp\left(\frac{aV_{BE}}{kT}\right)\right]$$

$$I_s = \frac{q^2A^2n_i^2\langle D_n\rangle}{Q_B}$$

$$\langle D_n\rangle = \frac{qA\int_0^{x_B} p(x)\mathrm{d}x}{\int_0^{x_B}\frac{p(x)}{D_n}\mathrm{d}x}$$

and

$$Q_B = qA\int_0^{x_B} p(x)\mathrm{d}x$$

The base current can be expressed in a similar form where current is a function of the excess charge in the base divided by the recombination time

$$Q'_B = qA\int_0^{x_B}[n(x) - n_0]\mathrm{d}x$$

$$I_{rB} = \frac{Q'_B}{\tau_n}$$

In the case of a MOSFET, the base region is the *p*-substrate region. In the base, it can be assumed the doping concentration in the lateral transistor structure is a constant. This simplifies the collector current equation [23].

$$I_C = I_s\left[\exp\left(\frac{qV_{BC}}{kT}\right) - \exp\left(\frac{aV_{BE}}{kT}\right)\right]$$

$$I_s = \frac{q^2A^2n_i^2\langle D_n\rangle}{Q_B} \simeq \frac{qA_en_i^2D_n}{N_aL_{eff}}$$

where A_e is the emitter area, D_n is the electron diffusion coefficient in the channel region, N_a is the doping concentration in the channel region, L_{eff} is the lateral bipolar base width of the lateral *npn* which is the effective channel length. For a MOSFET, the emitter area is equal to the product of the MOSFET channel width and the source/drain implant junction depth $W_{eff}X_J$. The effective area is the area which participates in the forward bias injection into the substrate.

$$I_s \simeq \frac{q(W_{eff}x_j)n_i^2D_n}{N_aL_{eff}} \equiv q\frac{W_{eff}}{L_{eff}}\frac{n_i^2D_n}{N_a}$$

For the base current, the equation can be expressed as

$$I_B = I_{0e}\left[\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right]$$

where the base current term I_{0e} is expressed as

$$I_{0e} \simeq \frac{qA_en_i^2D_p}{N_{DE}L_{pE}} \equiv \frac{q(W_{eff}x_J)n_i^2D_p}{N_{DE}L_{pE}}$$

The bipolar current gain is obtained from the ratio of the collector current to the base current. In the case of forward active voltage, the exponential term for the collector–base is negligible, and the bipolar current gain can be expressed solely as a function of the ratio of the two current expressions.

$$\beta_f = \frac{I_C}{I_B} \simeq \frac{I_s}{I_{0e}}$$

This can be expressed as

$$\beta_f = \frac{I_C}{I_B} \simeq \frac{I_s}{I_{0e}} \equiv \frac{D_n}{D_p}\frac{N_{DE}}{N_A}\frac{L_{pe}}{L_{eff}}$$

From the current model, avalanche generation can be expressed as a function of the multiplication factor and the total current flowing through the high electric field region. We can express the generation current as [74–81]

$$I_G = (M-1)(I_{DS} + I_C)$$

When the electric fields are low, M is unity, leading to no generation. As the electric field increases in the transistor, avalanche multiplication increases leading to the generation term. In this expression, some of the hole generation current will serve as base current for the lateral bipolar element, while some of the holes generated will enter the substrate serving as substrate current. The substrate current can be expressed as

$$I_{sx} = (M-1)(I_{DS} + I_C) - I_B$$

The substrate current in a MOSFET is typically an issue for hot electron generation and total power consumption. For ESD analysis, the substrate current is important because it is associated with the voltage drop that occurs locally in the MOSFET region at high currents.

$$V_{sx} = I_{sx} R_{sub}$$

hence, we can express the voltage drop in the substrate as

$$V_{sx} = [(M-1)(I_{DS} + I_C) - I_B] R_{sub}$$

When the voltage drop equals the forward bias voltage of the p–n junction formed between the source and the substrate, MOSFET snapback occurs. We can express this as the condition

$$V_{BE} = V_{sx} = [(M-1)(I_{DS} + I_C) - I_B] R_{sub}$$

or a condition for MOSFET snapback occurs when

$$V_{BE} \geqslant [(M-1)(I_{DS} + I_C) - I_B] R_{sub}$$

From this development, it is important to quantify the substrate resistance in order to anticipate the onset of MOSFET snapback. In a multi-finger structure, the MOSFET can be represented as a plurality of parallel MOSFET devices where the gate and the drain are coupled to common nodes, and the sources are independent nodes. The MOSFET fingers are also coupled together through the substrate potential drops. This introduces a complex model for the substrate current and the substrate resistances. This can be addressed through coupled system of equations, matrices, or transfer resistance representations. In the next chapter, substrate electrical and thermal models will be discussed.

3.6.2 Avalanche Breakdown and Snapback

In MOSFET devices, avalanche breakdown occurs in the metallurgical junction formed between the MOSFET drain and its supporting structure. In the case of an n-channel MOSFET, avalanche breakdown occurs in the $n+$ source/drain implant to $p-$ epitaxy (or $p-$ substrate) metallurgical junction. From the Townsend criteria

$$\int \alpha \mathrm{d}x = 1$$

where the integration of the impact ionization over the physical space where there exists a nonzero electric field. From this form, the avalanche multiplication factor M can be related to the impact ionization coefficient integrated over the depletion width and can be expressed as

$$M = \frac{1}{1 - \int \alpha \mathrm{d}x}$$

In a depletion region, the peak electric field is maximum at the center region of the dipole. In the analysis of the MOSFET, the model is simplified to express it as a function of the depletion width. Hence it can be integrated over the integral and stated in the form

$$M = \frac{1}{1 - \alpha x_{\mathrm{d}}}$$

From semiconductor physics, it is known that the depletion width can be expressed as a power of the applied voltage

$$x_{\mathrm{d}} \propto (V_{\mathrm{D}})^n$$

The power of this relationship is a function of the doping profile at the metallurgical junction. In this form, it is also clear that, when the voltage is greater than the avalanche breakdown voltage, the multiplication factor should increase rapidly. Hence heuristically, it is clear that the multiplication expression should satisfy the form

$$M = \frac{1}{1 - \left(\dfrac{V_{\mathrm{D}}}{V_{\mathrm{av}}}\right)^n}$$

In this form, as the drain voltage approaches the avalanche breakdown voltage, the multiplication factor approaches infinity. Amerasekera pointed out that the above relationship does not address the effect of the gate voltage on the electric field in a MOSFET gated diode region [74–80]. As the gate electrode voltage in a MOSFET is increased, the electric field in the drain region increases. From the expression

$$M \simeq \frac{1}{1 - \alpha(E) x_{\mathrm{d}}} = \frac{1}{1 - \alpha_0 x_{\mathrm{d}} \exp\left\{-\dfrac{B}{E}\right\}}$$

Substituting in a voltage condition where the electric field is the voltage over a physical distance, where the voltage across the depletion region is the drain voltage minus the drain saturation voltage, and the distance is the depletion region,

$$M \simeq \frac{1}{1 - (\alpha_0 x_{\mathrm{d}}) \exp\left\{-\dfrac{B x_{\mathrm{d}}}{\{V_{\mathrm{D}} - V_{\mathrm{d_{sat}}}\}}\right\}}$$

where the drain saturation velocity is expressed as

$$V_{d_{sat}} = \frac{V_G - V_t}{a + b(V_G - V_t)}$$

where saturation velocity is the voltage drive divided by a two-parameter expression in the denominator, and the voltage drive is the gate voltage minus the MOSFET threshold voltage.

The avalanche generation current can then be calculated from the multiplication factor and the current flowing through the drain junction. The current flowing through the drain structure is the MOSFET current flowing through the surface region (e.g. MOSFET source-to-drain current) as well as the current flowing from the parasitic bipolar transistor formed from the MOSFET source, epitaxy region, and the MOSFET drain forming a lateral *npn* transistor. In this case, the MOSFET drain and source serve as the bipolar junction transistor collector and emitter, and the epitaxial region serves as a base. The avalanche current can be expressed as

$$I_{av} = (M - 1)I = (M - 1)\{I_{DS} + I_C\}$$

The avalanche generation current flows to the *p*-substrate region of the *n*-channel MOSFET structure. A portion of this current flows to the base of the lateral parasitic *npn* transistor, serving as the emitter–base current, while the rest of the current will flow to the substrate region, as substrate current. Hence the avalanche current can be defined as the sum of the substrate current (flowing to the substrate contact), and the base current of the lateral *npn* transistor (serving as base drive current),

$$I_{av} = I_{sx} + I_B$$

Then the substrate current, can be estimated as

$$I_{sx} = I_{av} - I_B = (M - 1)\{I_{DS} + I_C\} - I_B$$

hence we can express the avalanche current, and the substrate current as a function of the impact ionization, the depletion width, drain voltage, drain saturation voltage, and MOSFET *n*-channel and parasitic *npn* current.

$$I_{av} \simeq \left(\frac{1}{1 - (\alpha_0 x_d) \exp\left\{ - \dfrac{B x_d}{\{V_D - V_{d_{sat}}\}} \right\}} - 1 \right) \{I_{DS} + I_C\}$$

and

$$I_{sx} \simeq \left(\frac{1}{1 - (\alpha_0 x_d) \exp\left\{ - \dfrac{B x_d}{\{V_D - V_{d_{sat}}\}} \right\}} - 1 \right) \{I_{DS} + I_C\} - I_B$$

3.6.3 Instability and Current Constriction Model

Spatial instability in MOSFET structures is a key source of ESD failures in semiconductor components [60–80]. MOSFET current constriction occurs in the MOSFET channel region under the MOSFET gate structure. MOSFET instability is a function of the interrelation of the electric potential and temperature gradient. During an ESD event, the electric potential under the MOSFET channel region and in the MOSFET channel region can influence the MOSFET failure. From the relationship between electric potential, maximum temperature, electrical resistivity and thermal conductivity, a set of contours can be created as a function of a fixed maximum temperature. The maximum temperature in the electrical constriction can be between the intrinsic temperature and the melting temperature. From our earlier development on spatial instability, and utilization of the spreading resistance relationship we can solve for the steady state constriction in a MOSFET channel region. From spreading resistance theory, the relationship between contact radius a and a contour at a distance d can be shown to satisfy a model

$$R_{\mathrm{s}} = \left\{\frac{1}{2\pi a}\right\} \tan^{-1}\left\{\frac{d}{a}\right\}$$

Then we have the relationship of the current restriction as

$$\left\{\frac{1}{2\pi a}\right\} \tan^{-1}\left\{\frac{d}{a}\right\} = \left\{\frac{1}{\mathrm{I_m}}\right\} \int_{T_{\max}}^{T} \mathrm{d}T \left[\frac{\rho K}{\sqrt{2\left\{\int_{T_{\max}}^{T} \rho(T')K(T')\,\mathrm{d}T'\right\}}}\right]$$

In MOSFET structures where current instability is evident, the failure analysis typically shows that the MOSFET region where the silicon is molten is significantly smaller than the width of the channel. Let that dimension $d = W$, and the parameter W_{eff} is the constriction region. Then

$$\left\{\frac{1}{2\pi W_{\mathrm{eff}}}\right\} \tan^{-1}\left\{\frac{W}{W_{\mathrm{eff}}}\right\} = \left\{\frac{1}{\mathrm{I_m}}\right\} \int_{T_{\max}}^{T} \mathrm{d}T \left[\frac{\rho K}{\sqrt{2\left\{\int_{T_{\max}}^{T} \rho(T')K(T')\,\mathrm{d}T'\right\}}}\right]$$

Since the region of current constriction is significantly smaller that the MOSFET channel width in an unballasted element, the above expression can be approximated as

$$\left\{\frac{1}{2\pi W_{\mathrm{eff}}}\right\} \tan^{-1}\left\{\frac{W}{W_{\mathrm{eff}}}\right\} \approx \frac{1}{4W_{\mathrm{eff}}} = \left\{\frac{1}{\mathrm{I_m}}\right\} \int_{T_{\max}}^{T} \mathrm{d}T \left[\frac{\rho K}{\sqrt{2\left\{\int_{T_{\max}}^{T} \rho(T')K(T')\,\mathrm{d}T'\right\}}}\right]$$

Then the effective width of the current constriction is

$$W_{\text{eff}} = \frac{\mathrm{I_m}}{\int_{T_{\max}}^{T} \mathrm{d}T \left[\dfrac{4\rho K}{\sqrt{2\left\{ \int_{T_{\max}}^{T} \rho(T')K(T')\,\mathrm{d}T' \right\}}} \right]}$$

In this form, it can be seen that the effective width is a function of the maximum current through the constriction region. In a multi-finger MOSFET structure, the total current will be divided by the number of fingers which are undergoing current constriction. If the substrate potential is well grounded, multiple current constrictions are observed in parallel. If the substrate potential is not well grounded, or is floating during an ESD test condition, the number of current constrictions in the multi-finger MOSFET are lower.

Figure 3.14 is an atomic force microscope (AFM) image of second breakdown in series cascode MOSFET devices. This is the first image of an integrated MOSFET where there are two in series. From the observation, the current constricts under the first

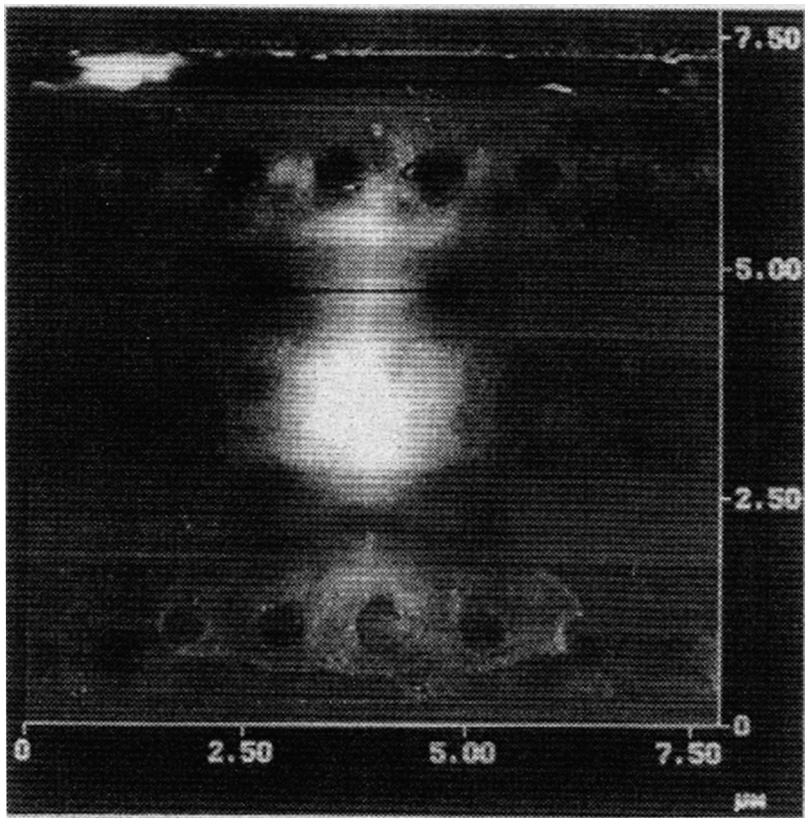

Figure 3.14 Atomic force microscope (AFM) image of second breakdown in series cascode MOSFET devices

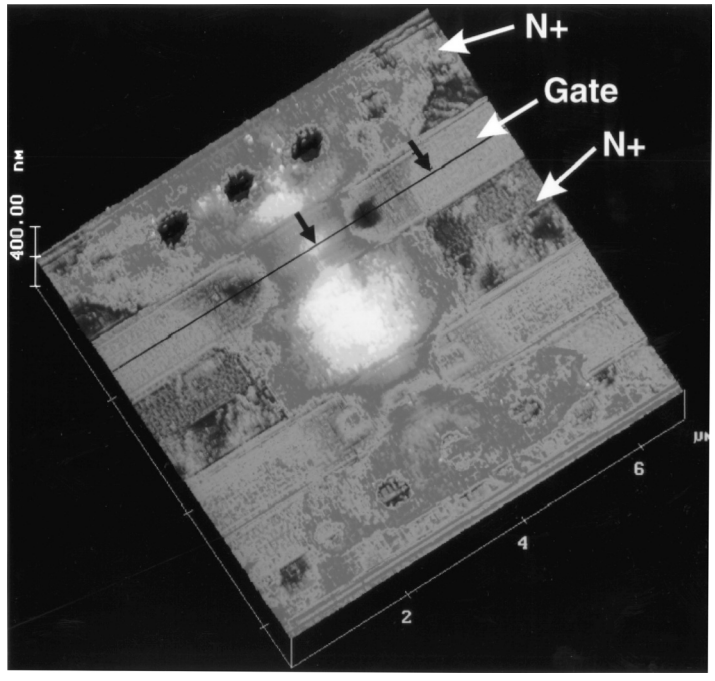

Figure 3.15 Atomic force microscope (AFM) image of second breakdown in series cascode MOSFET devices

MOSFET gate region, broadens out in the diffusion due to the low resistance and then reconstricts under the gate region.

Figure 3.15 is the atomic force microscope (AFM) image of second breakdown in series cascode MOSFET devices from a different perspective. From the image, it can be observed that there are two hourglass domains in the physical failure. Figure 3.16 is a cross-section cut through the MOSFET second breakdown across the gate region. Note that the symmetrical center of the constriction is the region of highest formation of molten silicon, and the region of peak heating.

3.6.4 Dielectric Breakdown

In MOSFET devices during ESD events, a key problem is dielectric breakdown of the MOSFET gate structure. Dielectric breakdown occurs typically from HBM, MM and CDM-type events [82–90].

In an HBM or MM event, the voltage occurs from a discharge on an input pad. As the voltage increases on the pad voltage and the MOSFET gate structure (e.g. in a receiver network). For a positive voltage, an *n*-channel MOSFET gate-to-substrate voltage increases, leading to a voltage stress between the MOSFET gate structure. For a negative HBM voltage, a negative MOSFET gate-to-drain voltage occurs. During the positive transition of an oscillatory event, such as the MM event, the MOSFET gate-to-substrate voltage is positive in the positive transition, and reverses in the

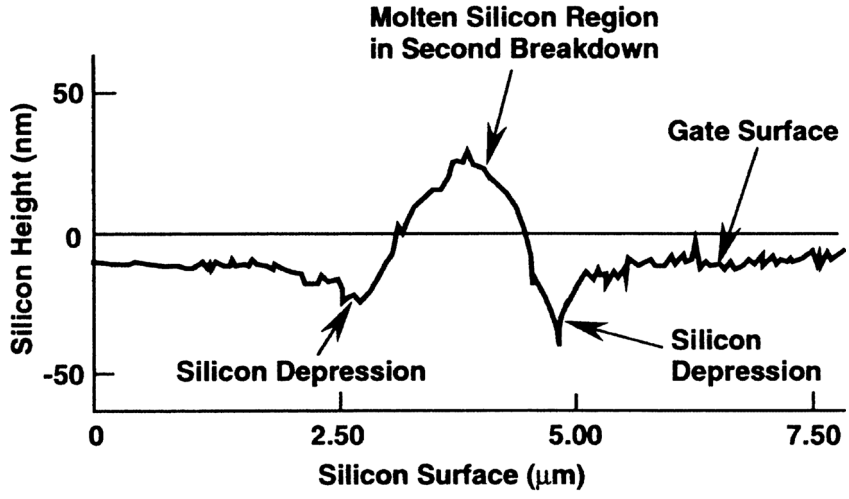

Figure 3.16 Atomic force microscope (AFM) image topography cut through the MOSFET second breakdown region

negative transition. In a typical MOSFET receiver network, the *n*-channel MOSFET gate fails during a positive HBM stress and the *p*-channel MOSFET fails during a negative voltage transition. In this failure mechanism, MOSFET receiver gate structures can be observed to fail in the polysilicon MOSFET gate structure at the connection into the device.

In a CDM event, the charge is formed in the substrate region. When the input pad is grounded a MOSFET gate-to-substrate voltage is established in the MOSFET channel region. The displacement current flows through the gate structure, leading to MOSFET gate failure. In these events, the failure mechanism typically is referred to as a MOSFET gate 'pinhole.' CDM events occur in *p*-channel MOSFETs when the power supply or well regions are charged and the *p*-channel MOSFET gate is grounded.

In the Fong and Hu model [82], the time to breakdown relationship can be represented as a function of two physical constants

$$t_{\mathrm{BD}} = C_1 \exp\{C_2/E_{0\mathrm{x}}\}$$

From this, a failure criteria can be established using this form where failure occurs when the integral over time is such that it is greater than the parameter C_1

$$\int \mathrm{d}t \exp\{-(C_2/E_{0\mathrm{x}}(t))\} > C_1$$

MOSFET gate dielectric failures also occur between the MOSFET gate-to-drain or the MOSFET gate-to-source regions. This is evident in both bulk and silicon on insulator (SOI) technology. In this case, latent damage is evident in the MOSFET gate-to-drain region which is observable from the MOSFET GIDL I-V shift or MOSFET overstress damage. Dielectric models will be further discussed in Chapter 8.

3.6.5 Gate Induced Drain Leakage (GIDL)

In MOSFET devices, prior to avalanche breakdown, electric field enhanced leakage mechanisms are evident in both *p*-channel MOSFETs, and *n*-channel MOSFETs [29,30,91,92]. In a MOSFET, as the gate-to-drain voltage increases, high electric fields occur in the MOSFET spacer and doped drain region and the metallurgical junction at the drain-to-substrate is modified and penetrates into the spacer and drain region along the device surface.

As the electric field increases, a depletion region is formed in the drain implant. At low electric fields, the leakage current is dominated by Shockley–Read–Hall generation. As the electric field is increased, the barrier height of the traps is modified by the electric field lowering the potential barrier. As the carrier height is reduced, the minority carriers flow from the traps to the drain-to-substrate metallurgical junction, leading to an electric-field-enhanced leakage mechanism. This Frenkel–Poole mechanism is more prevalent in *p*-channel MOSFET devices than observed in *n*-channel MOSFETs. As the electric field is increased, trap-to-band tunneling generation will occur, followed by the indirect bandgap band-to-band tunneling mechanism. With a further increase in the drain-to-source voltage, avalanche breakdown will occur. These high electric field effects must be addressed in the high-current effects of the MOSFET current model given this stress mode is present in the biasing of the MOSFET device. The high electric field enhanced leakage equations in the prior section can be used to quantify the effects in the MOSFET structure.

ESD events leading to stressing of the MOSFET gate-to-drain region can lead to hot carrier injection into the MOSFET drain/spacer region. During this process, the MOSFET drain-to-gate voltage will be modified, leading to a shifting of the electric field in the MOSFET drain-to-gate region. In observation of the MOSFET GIDL mechanism, shifts in the MOSFET GIDL *I–V* characteristics can be modified as a result of the electrical overstress. The shifting of the MOSFET GIDL I-V characteristic is known as MOSFET GIDL 'walkout.'

In the next chapters, we will start looking at the physical regions of semiconductor devices region by region. In this way, we can deconvolve the influence of the physical region and the corresponding processes. Chapter 4 focuses on substrates.

PROBLEMS

3.1. Derive the trigger criteria for a SCR for an *n*-substrate wafer as a function of the bipolar current gain of the *pnp* and *npn* devices.
3.2. Derive the trigger criteria for an SCR for an *n*-substrate wafer as a function of the collector-to-emitter transport factors, α_{npn} and α_{pnp}
3.3. Derive the *pnpn* SCR holding current for the low-voltage, high-current state, where there is no emitter–base shunting resistance.
3.4. Derive the *pnpn* SCR holding current for the low-voltage, high-current state, with emitter–base shunting resistance.
3.5. Given the SCR holding current I_H, formulate the sensitivity parameters $S^H{}_x$ where

$$S^H{}_x \equiv \frac{x}{I_H}\frac{\partial I_H}{\partial x}$$

and x are the holding current variables. Define the sensitivity parameters as a function of the bipolar current gains, well resistance and substrate resistance.

3.6. Derive the sensitivity parameters from the holding current relationship without the emitter resistances.

3.7. Derive the sensitivity parameter for the emitter resistances. Derive the sensitivity parameters as a function of all the holding current relationship which includes the emitter resistance terms.

REFERENCES

1. G. Wachutka. Rigourous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Transactions on Computer Aided Design*, 1990; **CAD-9**: 1141–1149.
2. H. S. Carslaw and J. C. Jaeger. *Conduction of Heat in Solids*, 2nd ed, London: Oxford University Press, 1959.
3. V. S. Alwin, D. Navon and L. Turgeon. Time dependent carrier flow in a transistor structure under nonisothermal conditions. *IEEE Transactions on Electron Devices*, 1977; **ED-24**: 1297–1303.
4. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*, New York: Springer-Verlag, 1984.
5. C. Jacoboni, C. Canali, G. Ottaviani and A. Quaranata. A review of some charge transport properties of silicon. *Solid State Electronics*, 1977; 20: pp. 77–89.
6. S. Voldman, S. Furkay and J. Slinkman. Three Dimensional Transient Electrothermal Simulation of Electrostatic Discharge Protection Circuits. *Proceedings of the EOS/ESD Symposium*, 1994; 246.
7. A. Amerasekera and C. Duvvury. *ESD in Silicon Integrated Circuits*. J. Chichester. Wiley, 1995.
8. S. Voldman, V. P. Gross, H. J. Hargrove, J. H. Never, J. A. Slinkman, M. O'Boyle, T. H. Scott and J. Delecki. Shallow trench isolation double-diode electrostatic discharge circuit and interaction with DRAM output circuitry. *Proceedings of the EOS/ESD Symposium*, 1992; 277.
9. S. Voldman and V. Gross. Scaling, optimization, and design considerations of electrostatic discharge protection circuits in CMOS technology. *Journal of Electrostatics*, 1994; **33** (3): 327–357.
10. S. Voldman. ESD protection in a mixed voltage and multi-rail disconnected power grid environment in a 0.50 and 0.25 μm channel length CMOS technologies. *Proceedings of the EOS/ESD Symposium*, 1994, 246.
11. S. Voldman and G. Gerosa. Mixed voltage interface ESD protection circuits for advanced microprocessors in shallow trench and LOCOS isolation CMOS technologies. *International Electron Device Meeting (IEDM) Technical Digest*, 1994; 227.
12. S. Voldman, G. Gerosa, V. P. Gross, N. Dickson, S. Furkay and J. Slinkman. Analysis of a snubber-clamped diode string mixed voltage interface ESD protection network for advanced microprocessor. *Proceedings of the EOS/ESD* Symposium, 1994; 125.
13. J. Never and S. Voldman. ESD failure mechanisms of shallow trench isolated ESD structures. *Proceedings of the EOS/ESD Symposium*, 1995; 273–288.
14. S. Voldman *et al*. Semiconductor process and structural optimization of shallow trench isolation-defined and polysilicon-bound source/drain diodes for ESD networks. *Proceedings of the EOS/ESD Symposium*, 1998; 151–160.
15. C. Richier *et al*. Investigation on different ESD protection strategies devoted to 3.3 V RF applications (2 GHz) in a 0.18 μm CMOS process. *Proceedings of the EOS/ESD Symposium*, 2000; 251–260.

16. S. Voldman. The effect of deep trench isolation, trench isolation, and sub-collector doping concentration on the ESD robustness of RF ESD diode structures in a BiCMOS SiGe Technology. *Proceedings of the EOS/ESD Symposium*, 2003; 214–223.
17. E. Worley and Bakulin. Optimization of input protection for high speed applications. *Proceedings of the EOS/ESD Symposium*, 2002; 62–73.
18. S. Voldman and L. R. Lanzerotti. The influence of process and design of subcollectors on the ESD robustness of ESD structures and silicon germanium heterojunction bipolar transistors in a BICMOS SiGe technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 347–356.
19. S. Voldman, R. Schulz, J. Howard, V. Gross, S. Wu, A.Yapsir, D. Sadana, H. Hovel, J. Walker, F. Assaderaghi, B. Chen, J. Y-C. Sun and G. Shahidi. CMOS-on-SOI ESD protection networks. *Proceedings of the EOS/ESD Symposium*, 1993; 221.
20. S. Voldman. Electrostatic discharge protection of silicon on insulator (SOI) technology. *IEEE International SOI Conference Symposium Proceedings, Session 5.1*, 1999; 68–72.
21. Y. Wang, P. Juliano, S. Joshi and E. Rosenbaum. Electro-thermal modeling of ESD diodes in bulk-Si and SOI technologies. *Proceedings of the EOS/ESD Symposium*, 2000; 430–436.
22. B. Streetman. *Solid State Electronic Devices*, 4th ed., Englewood Cliffs: Prentice Hall, 1995.
23. R. S. Muller and T. I. Kamins. *Device Electronics for Integrated Circuits*. Wiley 1977.
24. W. Schockley and W. T. Read. Statistics of recombination of holes and electrons. *Physical Review* 1952; **87** (5): 835–842.
25. R. N. Hall. Electron–hole recombination in germanium. *Physical Review* 1952; **87**: 387.
26. P. T. Landsberg. The band-to-band Auger effect in semiconductors. *Solid State Electronics* 1987; **30** (11): 1107–1115.
27. J. Dziewior and W. Schmid. Auger coefficients for highly doped and highly excited silicon. *Applied Physics Letters* 1977; **31** (5): 346–347.
28. J. D. Beck and R. Conradt. Auger recombination in silicon. *Solid State Communications* 1973; **13**: 93.
29. S. Voldman. Unified generation model for the analysis of electric field dependent leakage mechanisms in silicon devices. *PhD Thesis*, University of Vermont, 1991.
30. S. Voldman, J. B. Johnson, T. D. Linton and S.L. Titcomb. Unified generation model with donor and acceptor trap states for heavily doped silicon. *International Electron Device Meeting (IEDM) Technical Digest*, 1990; 349–352.
31. J. Frenkel. On pre-breakdown phenomena in insulators and electronic semiconductors. *Physical Review* 1938; **54**: 647–648.
32. J. L. Hartke. The three-dimensional Frenkel–Poole effect. *Journal of Applied Physics* 1968; **39**: 4871–4873.
33. A. K. Jonscher. Electronic properties of amorphous dielectric films. *Thin Solid Films*, 1967; **1**: 213–234.
34. R. H. Fowler and L. Nordheim. Electron emission in intense electric fields. *Proceedings of the Royal Society (London)* 1928; **121**: 626.
35. C. Zener. A theory of electrical breakdown of solid dielectrics. *Proceedings of the Royal Society (London)* 1933; **A145**: 523–529.
36. K. Mcafee, E. Ryder, W. Shockley and M. Sparks. Observations of Zener current in germanium *p–n* junctions. *Physical Review* 1951; **83**, 650–651.
37. L. D. Keldysh. Behavior of non-metallic crystals in strong electric fields. *Soviet JETP* 1958; **6**, (4): 763.
38. E. O. Kane. Zener tunneling in semiconductors. *Journal of Physics and Chemistry of Solids*, 1959; **12**, 181–188.
39. E. O. Kane. Theory of tunneling. *Journal of Applied Physics*, 1961; **32**, (1) 83–91.
40. P. J. Price and J. M. Radcliffe. Esaki tunneling. *IBM Journal of Research and Development*, October 1959; 364–371.

41. P. A. Wolff. Theory of electron multiplication in silicon and germanium. *Physical Review* 1954; **95**, (6): 1415–1420.
42. A. G. Chynoweth. Ionization rates for electrons and holes in silicon. *Physical Review* 1958; **109** (5): 1537–1540.
43. J. L. Moll and R. Van Overstraeten. Charge multiplication in silicon *p–n* Junctions. *Solid State Electronics* 1963; **6**: 147–157.
44. C. T. Kirk. A theory of transistor cutoff frequency (Ft) falloff at high current densities. *IEEE Transactions on Electron Devices* 1962; **ED-9**: 164–174.
45. E. O. Johnson. Physical limitations on frequency and power parameters of transistors. *RCA Review*, June 1965; 163–177.
46. J. M. Early. Maximum rapidly switchable power density in junction triodes. *IEEE Transactions on Electron Devices* 1959; **ED-6**: 322.
47. J. M. Early. Speed in semiconductor devices. *IRE National Convention*, March 1962.
48. J. M. Early. Structure determined gain–band product of junction triode transistors. *Proceedings of the IRE*, December 1958.
49. J. M. Goldey and R. M. Ryder. Are transistors approaching their maximum capabilities. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, 1963; 20.
50. R. L. Pritchard. Frequency response of grounded base and grounded emitter Junction Transistor. *IEEE Winter Meeting*, January 1954.
51. W. Schockley. Theories of high values of alpha for collector contacts on germanium. *Physical Review*, 1950; **78**: 294–295.
52. J. J. Ebers. Four terminal *pnpn* transistors. *Proceedings of the IRE*, 1952; **40**: 1361–1364.
53. J. L. Moll, M. Tannnenbaum, J. M. Goldey and N. Holonyak. *pnpn* transistor switches. *Proceedings of the IRE* 1956; **44**: 1174–1182.
54. F. E. Gentry *et al*. *Semiconductor Controlled Rectifiers*. Englewood Cliffs: Prentice-Hall, 1964.
55. B. L. Gregory and B. D. Schafer. Latchup in CMOS Integrated Circuits. *IEEE Transactions on Nuclear Science*, 1973; **NS-20**: 293–299.
56. D. B. Estreich. The physics and modelling of latchup in CMOS integrated circuits. *PhD Thesis*, Stanford University, 1981.
57. R. R. Troutman, *Latchup in CMOS Technology: The Problem and the Cure*. Kluwer, 1984.
58. V. Puvvuda, V. Srinivasan and V. Gupta. A scalable analytical model for ESD *n*-well resistors. *Proceedings of the EOS/ESD Symposium*, 2000; 437–446.
59. P. L. Hower and V. G. K. Reddi. Avalanche injection and second breakdown in transistors. *IEEE Transactions on Electron Devices* 1970; **ED-17**: 320–335.
60. H. A. Schafft and J. C. French. Second breakdown in transistors. *IEEE Transactions of Electron Devices* 1962; **ED-9**: 129–136.
61. R. M. Scarlett, W. Schockley and R. H. Haitz. Thermal instabilities and hot spots in junction transistors. In: M. F. Goldberg and J. Vaccaro (eds), *Physics of Failure in Electronics*: Baltimore, Spartan Books, 1963; 194–203.
62. R. J. Nienhuis. Second breakdown in forward and reverse base current region. *IEEE Transactions on Electron Devices* 1966; **ED-13**: 655–662.
63. H. B. Grutchfield and T. J. Moutoux. Current mode second breakdown in epitaxial planar Transistors. *IEEE Transactions on Electron Devices* 1966; **ED-13**: 743.
64. P. L. Hower and V. G. K. Reddi. Avalanche injection and second breakdown in transistors. *IEEE Transactions on Electron Devices* 1970; **ED-17**: 320–335.
65. D. M. Tasca, J. C. Peden, and J. Miletta. Non-destructive screening for thermal second breakdown. *IEEE Transactions on Nuclear Science* 1972; **NS-19**: 57–67.
66. R. W. Dutton. Bipolar transistor modeling of avalanche generation for computer circuit simulation. *IEEE Transactions on Electron Devices*, 1975; **ED-22**: 334–338.
67. A. Ward. An electrothermal model of second breakdown. *IEEE Transactions on Nuclear Science* 1976; **NS-23**: 1679–1684.

68. S. Gaur, D. Navon and R. Teerlinck. Transistor design and thermal stability. *IEEE Transactions on Electron Devices* 1976; **ED-20**: 527–534.
69. S. K. Ghandi. *Semiconductor Power Devices*. J. New York: Wiley, 1977.
70. K. Koyanagi, K. Hane and T. Suzuki. Boundary conditions between current mode and thermal mode second breakdown in epitaxial planar transistors. *IEEE Transactions on Electron Devices* 1977; **ED-24**: 672–678.
71. J. Yee, W. Orvis, L. Martin and J. Peterson. Modeling of current and thermal mode second breakdown phenomenon. *Proceedings of the EOS/ESD Symposium*, 1982; 76–81.
72. W. J. Orvis, C. F. McConaghy, J. H. Yee, G. H. Khanaka, L. C. Martin and D. H. Lair. Modeling and testing for second breakdown phenomena. *Proceedings of the EOS/ESD Symposium*, 1983; 108–117.
73. D. Scott, G. Giles and J. Hall. A lumped element model for simulation of ESD failures in silicided devices. *Proceedings of the EOS/ESD Symposium*, 1986; 41–47.
74. G. Krieger. The dynamics of electrostatic discharge prior to bipolar action related snapback. *Proceedings of the EOS/ESD Symposium*, 1989; 136–144.
75. M. Reisch. On bistable behavior and open-base breakdown of bipolar transistors in the avalanche regime–modeling and applications. *IEEE Transactions on Electron Devices* 1992; **ED-39**: 1398.
76. A. Amerasekera, L. V. Roozendaal, J. Bruins and F. Kuper. Characterization and modeling of second breakdown for the extraction of ESD-related process and design parameters. *IEEE Transactions on Electron Devices*, 1991; **ED-38**: 2161.
77. E. A. Amerasekera, W. Van den Abeelen, L. J. van Roozendahl, M. Hannemann, P. J. Schofield. ESD failure modes: characteristics, mechanisms, and process influences. *IEEE Transactions on Electron Devices*. 1992; **ED-39**: 2.
78. A. Amerasekera, M. C. Chang, J. Seitchik, A. Chatterjee, K. Mayaram and J. H. Chern. Self heating effects in basic semiconductors structures. *IEEE Transactions on Electron Devices* 1993; **ED-40**: 1836.
79. A. Amerasekera and J. Seitchik. Electrothermal behavior of deep submicron *n*MOS transistors under high current snapback ESD/EOS conditions. *International Electron Device Meeting (IEDM) Technical Digest*, 1994; 446–449.
80. A. Amerasekera and R. Chapman. Technology design for high current and ESD robustness in a deep submicron process. *IEEE Transactions on Electron Device Letters* 1994; **15**: 383–385.
81. S. Verdonckt-Vandebroek, S. S. Wong, J. Woo and P. K. Ko. High current gain lateral bipolar action in a MOSFET structure. *IEEE Transactions on Electron Devices* 1991; **ED-38**: 2487–2496.
82. Y.-P. Fong and C. Hu. The effects of high electric field transients on thin gate oxide MOSFETs. *Proceedings of the EOS/ESD Symposium*, 1987; 252–257.
83. D. L. Lin. ESD sensitivity and VLSI technology trends: thermal breakdown and dielectric breakdown. *Proceedings of the EOS/ESD Symposium*, 1993; 73–82.
84. A. Amerasekera and D. Campbell. ESD pulse and continuous voltage breakdown in MOS capacitors structures. *Proceedings of the EOS/ESD Symposium*, 1986; 208–213.
85. D. L. Lin and T. Welsher. From lightning to charged-device model electrostatic discharges. *Proceedings of the EOS/ESD Symposium*, 1992; 68–75.
86. M. Bridgewood and R. Kelly. Modeling the effects of narrow impulsive overstress on capacitive test structures. *Proceedings of the EOS/ESD Symposium*, 1985; 84–91.
87. M. J. Tuncliffe, V. M. Dwyer and D. S. Campbell. Experimental and theoretical Studies of EOS/ESD oxide breakdown in unprotected MOS structures. *Proceedings of the EOS/ESD Symposium*, 1990; 162–168.
88. Y. Kitamura, H. Kitamura, K. Nakanishi and Y. Shibuya. Breakdown of thin gate-oxide by application of nanosecond pulse as ESD test. *Proceedings of International Test and Failure Analysis (ITSFA)* 1989; 193–199.

89. T. J. Maloney. Designing MOS inputs and outputs to avoid oxide failure in the charged device model. *Proceedings of the EOS/ESD Symposium*, 1988; 220.
90. R. G. Renninger. Mechanism of charged device electrostatic discharge. *Proceedings of the EOS/ESD Symposium*, 1991; 127.
91. S. Voldman, J. Bracchita and D. Fitzgerald. Band-to-band tunneling and thermal gate-induced drain leakage. *Device Research Conference (DRC), Session IIA-8*, June 1988.
92. J. R. Spehar, R. A. Colclaser and C. B. Fleddermann. The effect of oxidation of the poly gate on the ESD performance of CMOS ICs. *Proceedings of the EOS/ESD Symposium*, 1994; 257.

4 Substrates and ESD

4.1 METHODS OF SUBSTRATE ANALYSIS

In this chapter, the influence of the substrate on ESD protection is addressed from a general approach. The substrate influences diodes, MOSFETs, bipolar transistors, and silicon-controlled rectifiers. Substrate electrical and thermal modeling influences all circuits and ESD elements. The general concepts and models can be applied to the diode, the bipolar transistor, the MOSFET, and the silicon-controlled rectifier. Teaching from the generalist approach, the substrate is a semi-infinite domain containing a plurality of devices, which can be treated independently or coupled. Substrate model concepts such as semi-infinite domain models, to transmission line models, and lossy transmission lines, are discussed. This is followed by electrical and thermal models using the Green's function methods and new geometrical models. Additionally, to address multi-finger or a plurality of elements, the transfer resistance concept, used by Ron Troutman for latchup, is shown for more complicated scenarios and coupling phenomena. These models can be used for ESD, noise or latchup. Another approach is viewing the substrate as a stratified media. This technique and the flux-potential transfer matrix approach are used widely at M.I.T. to address problems in continuum electromechanics, and in electrodynamics.

These methods are applicable for problem solving in the ESD area for variable doped regions, SOI wafers to GaAs substrates. The substrate and substrate modeling also play a key role in understanding silicon-on-insulator (SOI) devices. In the electrostatic discharge (ESD) discipline considerable energy is invested in evaluating the substrate effects at high current. The substrate also serves as a means of providing coupling within an ESD structure, or between different elements within the ESD circuit. As a result, evaluating from a general approach will provide some synergy to the different devices, as well as related matters such as noise coupling, and latchup, which are synergistically connected to ESD issues as well.

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

4.2 SUBSTRATE AS A SEMI-INFINITE DOMAIN

For the case of the semi-infinite region (Figure 4.1), we can modify the solution given by the parallelepiped in a infinite region, by defining a boundary condition and applying the method of images [1].

Analogous to the electrostatic problem, defining the temperature T at a point P in an infinite space from a point heat source q at point A, and an point heat sink $-q$ at point A', there exists a locus of points where the temperature is zero. This plane is the plane which bisects the line AA' at right angles. Hence an 'image' of opposite polarity and strength for an infinite medium exists on the opposite side of the plane. The temperature at the plane formed by the points P is equal to zero, is and satisfies the Laplace equation in the semi-infinite half-space except at the point heat source q at point A.

It is also known that any temperature field produced by any number of point heat sources, that we can select a constant temperature contour (e.g. isotherm) and replace it with a thermal conductor. The heat sources on either side of the isotherm are the images of the other side.

For the case of the adiabatic boundary condition, we need to require that the derivative of the temperature is zero at the boundary condition. Assume the parallel piped is displaced below an infinite plane at $z = 0$ by distance D. Let the boundary condition at the $z = 0$ plane be an adiabatic boundary condition. This form then assumes that the heat flux for all times $t > 0$ is zero for all points in the plane. Given that we are interested in the temperature field in the semi-infinite space below the plane $z = 0$, an image source that exist on the upper half of the infinite space can be used to provide the boundary condition of interest in the semi-infinite lower half-space.

Using a parallelepiped of identical dimensions displaced a distance D above the $z = 0$ plane, a plane of symmetry is established that has a solution where all points along the plane $z = 0$ have the heat flux at the surface equal to zero. To solve the problem of the semi-infinite domain, we can utilize the method of images by modification of the Green's function in the infinite domain so that it applies to the semi-infinite domain. From the time-dependent heat equation, with a constant thermal conductivity and constant specific heat–mass density product, the Laplacian of temperature is proportional to the partial derivative of temperature as a function of time. As in the problem of an infinite domain, we have

$$T(x, x'; y, y'; z, z'; t) = \int_{t'=0}^{t'=t} \mathrm{d}t' \frac{Q(t')}{8[\pi\kappa(t - t')]^{3/2}} \exp\left\{-\frac{r^2}{4\kappa(t - t')}\right\}$$

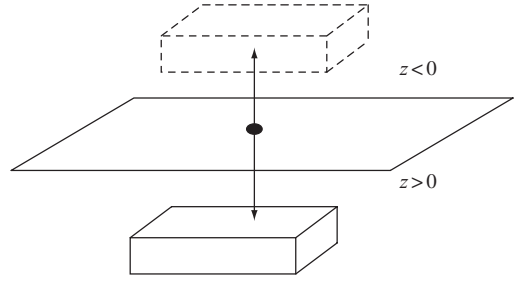

Figure 4.1 Semi-infinite domain and method of images

Expressing this as a function of power dissipated in an infinitesimal volume formed between x' and $x'+\mathrm{d}x'$, y' and $y+\mathrm{d}y'$, and z' and $z+\mathrm{d}z'$, and defining the power dissipated as normalized to the volume V,

$$T(x,x';y,y';z,z';t)=\frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t}\mathrm{d}t'\,\frac{\{P(t')\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'\}}{[(t-t')]^{3/2}}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}$$

To evaluate the temperature over all space from all infinitesimal volumes, the expression can be integrated over the infinite volume Ω. In the expression, the spatially independent terms can be removed from the integration over space, and the temperature can be expressed as

$$T(x,y,z;t)=\frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t}\mathrm{d}t'\,\frac{P(t')}{[(t-t')]^{3/2}}\int_{\Omega}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$$

For evaluation of a parallelepiped in an infinite medium let us assume a source has the dimension W in the x-direction, L in the y-direction, and H in the z-direction, but applying a parallelepiped displaced distance D below the boundary condition $z=0$ and an image source of equal and opposite strength above the $z=0$ plane at $z=D$

$$T(x,y,z;t)=\frac{1}{8\rho cV(\kappa)^{3/2}}\int_{t'=0}^{t'=t}{}'\,\frac{P(t')\mathrm{d}t'}{[(t-t')]^{3/2}}F(x-x',y-y',z-z',t-t')$$

with

$$F(x-x',y-y',z-z',t-t')=Fx(x-x':t-t')Fy(y-y';t-t')Fz(z-z';t-t')$$

where

$$F_x(x-x',t-t')=\int_{-W/2}^{W/2}\exp\left\{-\frac{(x-x')2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}x'}{\sqrt{\pi}}$$

$$F_y(y-y',t-t')=\int_{-L/2}^{L/2}\exp\left\{-\frac{(y-y')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}y'}{\sqrt{\pi}}$$

$$F_z(z-z',t-t')=\int_{-(D+H)}^{-D}\exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}z'}{\sqrt{\pi}}+\int_{D}^{D+H}\exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}z'}{\sqrt{\pi}}$$

The integral expression can be expressed as error functions using a transformation of variables. The expression for temperature in an infinite medium can be put in the form

$$T(x,y,z;t)=\frac{1}{8C}\int_{t'=0}^{t'=t}P(t')H(x,y,z;t-t')\mathrm{d}t'$$

where $V = LWH$ is the volulme, and letting $C = c\rho V$
where we can write the function containing the spatial dependence as the product of the error functions

$$H(x,y,z;t-t') = H(z;t-t') \prod_{i=x,y} \left[erf\left(\frac{(L_{xi}/2)+x_i}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{(L_{xi}/2)-x_i}{\sqrt{4\kappa(t-t')}} \right) \right]$$

and

$$H(z;t-t') = \left[erf\left(\frac{z+D+H}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{-D-z}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{z-D}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{D+H-z}{\sqrt{4\kappa(t-t')}} \right) \right]$$

4.3 SUBSTRATE AS A STRATIFIED MEDIUM USING THE TRANSFER MATRIX APPROACH

Substrate regions and semiconductors can consist of planar layers of conducting and nonconducting regions forming a stratified medium [2,3]. For example, substrates can consist of p^{++} wafers with p-epitaxial regions, p-wafers with a p^{++} heavily doped buried layer (HDBL), strained silicon with Si/SiGe/Si regions, and even silicon-on-insulator (SOI) forming a silicon/silicon dioxide/silicon wafer. In these problems, the electrical and thermal variables (e.g. thermal conductivity k) are constant in the region, but different from region to region. For many of these problems, there is little interest within the physical region of the thermal field, but there is interest in the temperatures and heat flux at the interfaces. For these above structures, the field–flux transfer relations are ideal. In the case of a volumetric region where no heat generation source is present, the Laplace equation is satisfied. The solution of the Laplace equation can be represented as a matrix representation in terms of the relationship between the heat flux and temperature at two interfaces. In this form, the solution can be solved using a 'transfer matrix' where the transfer matrix and its corresponding boundary condition satisfies the Laplace equation. This methodology can be used in planar, cylindrical and spherical geometric regions.

In the case of a region with boundaries, or material differences, a stratified medium can be represented as a set of n regions where the solutions at the interfaces are of interest, or only the relationship of the first and last outer boundaries. Multiple films of planar films can be represented as a stratified medium (Figure 4.2) where the material properties may vary in the different physical regions. Note that any physical regions with variable electrical or thermal properties can be segmented into a plurality of regions where an average electrical or thermal property is defined in that region.

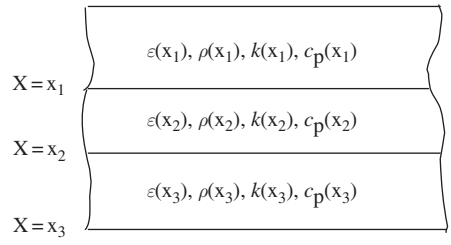

Figure 4.2 Substrate as a stratified medium

To evaluate the thermal flux and the temperature field, the system with uniform thermal properties satisfies the Laplace equation and the flux–potential relationship

$$q = -k\nabla T$$

Addressing the thermal problem, the relationship between the thermal flux from the thermal potential on the boundaries, let us find a matrix representation of the relationship of the forward heat flow from the boundary conditions of temperature. Conversely, if we know the flux on the boundaries, we can determine the potential on the boundaries. The Laplace equation can be transformed according to the method of separation of variables to address the multiple dimensional problem or one-dimensional transient thermal analysis. Taking the case of the transient thermal heat conduction, the temperature is assumed to be a function of space and time and can be expressed as

$$T(x, t) = T(x)\Gamma(t)$$

Substitution of the function into the heat equation, the form can be represented

$$\frac{1}{T(x)}\frac{\mathrm{d}^2 T(x)}{\mathrm{d}x^2} = \frac{1}{\alpha\Gamma}\frac{\mathrm{d}\Gamma}{\mathrm{d}t} = -\lambda^2$$

Hence, we can solve the spatial variation independently, by solving for the second-order form of the modified Laplace equation as

$$\frac{\mathrm{d}^2 T(x)}{\mathrm{d}x^2} + \lambda^2 T(x) = 0$$

Let us define a temperature and heat flux at a first boundary and second boundary. A solution of the equation by inspection, satisfies the form

$$T(x) = T^{\alpha}\frac{\sin(\lambda x)}{\sin(\lambda\Delta)} - T^{\beta}\frac{\sin\{\lambda(x-\Delta)\}}{\sin(\lambda\Delta)}$$

where the first surface β is at $x = 0$ and the second surface α is at $x = \Delta$.

The derivative of temperature is

$$\frac{\mathrm{d}}{\mathrm{d}x}T(x) = T^{\alpha}\frac{\cos(\lambda x)}{\sin(\lambda\Delta)} - T^{\beta}\frac{\cos\{\lambda(x-\Delta)\}}{\sin(\lambda\Delta)}$$

From this relationship, we can form a forward matrix relation of the derivative of temperature at the boundaries to the temperature at the boundaries [2]

$$\begin{bmatrix} \left.\frac{\mathrm{d}T}{\mathrm{d}x}\right|^{\alpha} \\ \left.\frac{\mathrm{d}T}{\mathrm{d}x}\right|^{\beta} \end{bmatrix} = \lambda \begin{bmatrix} \cot(\lambda\Delta) & \frac{-1}{\sin(\lambda\Delta)} \\ \frac{1}{\sin(\lambda\Delta)} & -\cot(\lambda\Delta) \end{bmatrix} \begin{bmatrix} T^{\alpha} \\ T^{\beta} \end{bmatrix}$$

From the relationship of heat flux to temperature we can represent this in the form

$$\begin{bmatrix} q^{\alpha} \\ q^{\beta} \end{bmatrix} = -\lambda k \begin{bmatrix} \cot(\lambda\Delta) & \frac{-1}{\sin(\lambda\Delta)} \\ \frac{1}{\sin(\lambda\Delta)} & -\cot(\lambda\Delta) \end{bmatrix} \begin{bmatrix} T^{\alpha} \\ T^{\beta} \end{bmatrix}$$

This form allows for the representation of the heat flux from the temperature at the boundaries. Representing the temperature as a function of the heat flux

$$\begin{bmatrix} T^{\alpha} \\ T^{\beta} \end{bmatrix} = -\frac{1}{\lambda k} \begin{bmatrix} \cot(\lambda\Delta) & \frac{-1}{\sin(\lambda\Delta)} \\ \frac{1}{\sin(\lambda\Delta)} & -\cot(\lambda\Delta) \end{bmatrix} \begin{bmatrix} q^{\alpha} \\ q^{\beta} \end{bmatrix}$$

From this matrix approach, a region can be represented as n regions of a uniform property. The solution to the stratified medium can be solved using matrix multiplication. From an electric circuit analogy, each physical film can be thought of as a two-port film where the information is zero at the terminals. This methodology can be used for the transient one-dimensional solution or oscillatory signals that are separable according to the form required for separation of variables.

Hence, viewing the substrate as a vertically stratified medium, the temperature and heat flux field at the interfaces can be obtained without reevaluation of the Laplace equation, and the evaluation is reduced to matrix multiplications. This method can be modified to the cylinder and spherical source form where imaginary boundaries can be used to determine the cylindrical and spherical heat fluence and temperature at the pseudo-interfaces.

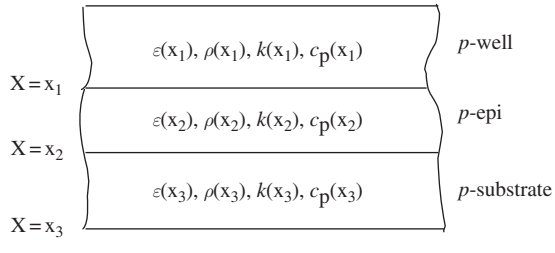

Figure 4.3 Epitaxial wafer as a stratified medium

Figure 4.3 is an example of the application of a stratified medium problem to an epitaxial wafer. In an epitaxial wafer, there is a *p*-well region, a *p*-epitaxial region and a *p*-substrate. In this case the region can be subdivided into regions where the temperature and thermal fluence can be evaluated as independent regions by evaluation of the conditions on the boundaries.

4.4 SUBSTRATE TRANSMISSION LINE MODEL

The substrate spans the physical space in a semiconductor chip, leading to both coupling effects as well as distribution effects. The modeling of the substrate is key to understanding the response to the circuit where the active region is influenced by the local potential. Two primary cases are of interest.

The first case is where the element is a plurality of active elements, where each spatial region has at least one node which are independent from the other active elements. In this case, a common electrical node may be present, but some of the nodes can be regarded as independent. For example, in a multi-finger MOSFET transistor, the gate structure is common, but the source or drain nodes may be electrically independent.

A second case of interest is where there are spatially separated regions with no active device between them. For example, a common electrical input may be connected to a first segment of an ESD device, and at a second point in space, there is a second segment of significant distance away wherein the second device is near a receiver network. This occurs in ESD networks with a first primary stage near a bond pad, and a second circuit near a receiver network within the semiconductor chip.

Substrate modeling is addressed by using lumped resistor elements under the semiconductor devices. In small ESD devices, typically, the substrate is modeled as a single resistor element. In multi-finger elements, the lateral resistance between the two physical elements can lead to different circuit responses at high current or injection into the substrate region. To address substrate resistance effects, a different lumped resistor element is placed under each independent active device region. The substrate is modeled as a lateral resistance to the electrical contact region, and vertical resistances under each physical device region.

As in the example of the stratified medium to evaluate the vertical thermal and electrical properties, the transmission line model is useful to evaluate the lateral physics. Figure 4.4 is an example of a vertical profile of a *p*-well, a *p*-epitaxial region and a p^+-substrate region.

A lateral model of discrete resistors can be formed to represent the substrate as a ladder network. Figure 4.5 demonstrates the representation as a resistor ladder network.

A better representation is a full transmission line representation. For high-current phenomena near the surface of the wafer, the substrate can be modeled as a lossy transmission line. A lossy transmission line model is suitable for current flow between two regions near the surface of the wafer and cases where current can flow from back-surface contacts [4–6]. From the general form of a lossy transmission line, we can express the current and voltage conditions as

$$V(y) = V_1 \exp\left\{-\frac{y}{\gamma}\right\} + V_2 \exp\left\{-\frac{L-y}{\gamma}\right\}$$

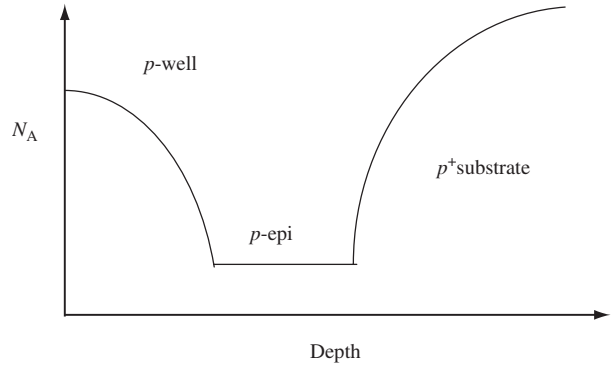

Figure 4.4 Cross sectional profile of a p-well/p-epi/p^+ substrate wafer

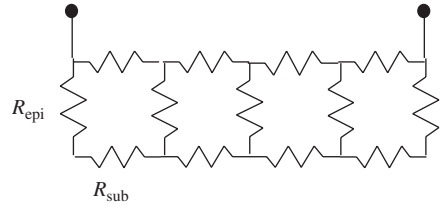

Figure 4.5 Discrete resistor ladder network representation of substrate

and

$$I(y)\frac{1}{Z_0}\left[V_1\exp\left\{-\frac{y}{\gamma}\right\}-V_2\exp\left\{-\frac{L-y}{\gamma}\right\}\right]$$

These current and voltage conditions represent a lossy transmission line with the transmission propagating in the y-direction, whose transmission line length is L with the loss factor, γ. This is shown in Figure 4.6.

Assuming a transmission line with a termination resistance R_1 at terminal 1 and a termination resistance R_2 at terminal 2, we can express the transmission line system in the form

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \frac{1}{Z_0}\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

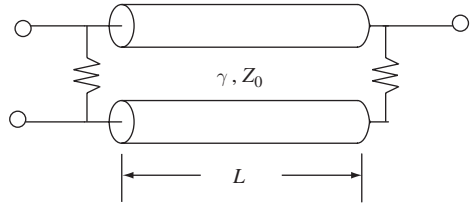

Figure 4.6 Lossy Transmission line (TL) model representation of the substrate

where the main diagonal terms are expressed as

$$C_{11} = (1 + Z_0/R_1)$$

$$C_{22} = (1 + Z_0/R_2)$$

$$C_{12} = (-1 + Z_0/R_1)\exp\left\{-\frac{L}{\gamma}\right\}$$

$$C_{21} = (-1 + Z_0/R_2)\exp\left\{-\frac{L}{\gamma}\right\}$$

Solving for the voltage equations at terminals 1 and 2, the matrix can be inverted and can be represented as a function of the two current conditions.

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta}\begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix}\begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$$

where the determinant is $\Delta = C_{11}C_{22} - C_{12}C_{21}$.

From this form, the voltage between two points in the substrate can be evaluated, and the input impedance can be solved as a function of the characteristic impedance, the reflection coefficient and the loss factor. The input impedance can be expressed as

$$Z_{\text{in}} = \frac{V(0)}{I(0)} = Z_0\frac{1 + \Gamma_{\text{R}}\exp\left\{-\frac{2L}{\gamma}\right\}}{1 + \Gamma_{\text{R}}\exp\left\{-\frac{2L}{\gamma}\right\}}$$

and the reflection coefficient can be expressed as

$$\Gamma_{\text{R}} = \frac{V_2}{\left[V_1\exp\left\{-\frac{L}{\gamma}\right\}\right]}$$

The transmission line case can be simplified by addressing the case that current is not flowing to terminal 2. This can represent the case of the current flowing to a second region such as the back surface of the wafer. In this case, the surface potential may be equal the transmission line voltage. This can be solved for from the matrix relationship, as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta}\begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix}\begin{bmatrix} I_1 \\ 0 \end{bmatrix}$$

where

$$\psi_s(y) = V(y) = V_1\exp\left\{-\frac{y}{\gamma}\right\}\left[1 + \Gamma_{\text{R}}\exp\left\{-\frac{2(L-y)}{\gamma}\right\}\right]$$

and

$$I(y) = \frac{1}{Z_0} V_1 \exp\left\{-\frac{y}{\gamma}\right\}\left[1 - \Gamma_R \exp\left\{-\frac{2(L-y)}{\gamma}\right\}\right]$$

and

$$V_1 = \frac{Z_0}{\Delta} C_{22} I_1$$

In the case of a MOSFET transistor, current is flowing from the device to the substrate contact during an ESD event. In the case of a ESD diode network, the current flows from a diode junction to the substrate contact. In the case where a single element is involved in the interaction, and only one substrate contact is local to the element, we can assume the current flowing from the substrate is flowing from the first terminal to a second terminal, where the first terminal is a contact, and the second is an element. If the physical device is wide, we can ignore current flowing in the z-direction of the transmission line. In these cases, the contact are on the top of the physical wafer.

For the case where the current flowing from the first terminal equals that of the second terminal, then $I_2 = -I_1$. The expressions for V_1 and V_2 coefficients can be expressed as a function of the current I_1, and the loss factors in the transmission line

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta} \begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix} \begin{bmatrix} I_1 \\ -I_1 \end{bmatrix}$$

and surface potential,

$$\psi_s(y) = V(y) - V(L) = \left(V_1 \exp\left\{-\frac{y}{\gamma}\right\} - V_2\right)\left[1 - \exp\left\{-\frac{2(L-y)}{\gamma}\right\}\right]$$

Generalizing this approach, we can assume that the current in one element is related to the second, but that they are not equal. This can be represented as a percentage of the first term. Hence a third derivation can assume that $I_2 = -\delta I_1$ where δ is a percentage of the current.

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta} \begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix} \begin{bmatrix} I_1 \\ -\delta I_1 \end{bmatrix}$$

To address the case where surface structures are present along the interface between the points of interest, the substrate can be represented as a plurality of transmission lines in series, where insertion resistors, or stubs and other elements can interrupt the transmission line terminations. The voltage and current conditions can use matrix multiplication to introduce the terminations and boundary conditions. For example, along the interface, there are n^+ implants and other diffusions which can modify the current and voltage characteristics along the surface. In this fashion, multiple elements can intersect the transmission line, to address more complicated circuit environments.

4.5 SUBSTRATE LOSSY TRANSMISSION LINE MODELS

The substrate transmission line model characteristics will be a function of the doping concentration profile in the substrate. The doping concentration profile will have a strong effect on the loss characteristics of the transmission line model [4–6].

The cases of interest primarily are the heavily doped substrate with a thin epitaxial region, and the lightly doped substrate.This structure has been implemented to improve CMOS latchup, and soft error rate (SER) concerns [7–13]. For the *p*-type heavily doped substrate, the cases of interest are the lightly doped epitaxy without a *p*-well implant, the *p*-well case with an epitaxial flat zone, and the case of the *p*-well with a connecting implant. For the the lightly doped substrate, the types of substrate models of interest are the *p*-well with *p*-wafer, the *p*-wafer with heavily doped buried layer, and the lightly doped wafer with the heavily doped buried layer and a termination implant. Additionally, this can be utilized for models of *n*-substrates and triple wells. In all cases, the transmission line representation must address the transmission line loss factor, the characteristic impedance, and the termination resistances.

Troutman showed that a transmission line representation can be determined if we regard the system as a conductive region near the device surface, a second conductive region below the surface, and two termination resistors between the two conductive regions. This forms the lossy transmission line with the termination resistances related to the 'up' resistance between the two physical regions.

For the lossy transmission line development, the loss factor and the characteristic impedance are important for the evaluation. These can be expressed as the series resistance and the shunt conductance. The loss factor can be expressed as

$$\gamma = \frac{1}{\sqrt{rg}}$$

and the characteristic impedance

$$Z_0 = \sqrt{\frac{r}{g}}$$

The series resistance per unit length can be evaluated as

$$r = \frac{1}{W} \frac{1}{\int_0^{x_1} \sigma(x) \mathrm{dx}}$$

where the integration is completed to an arbitrary depth where one initiates the transmission line conductive plate. For the case of the implant near the device surface, with a background doping of N_{epi}, we can define a concentration for the surface region as

$$N_{\text{A}}(x) = N_{\text{epi}} + N_{\text{A}} \exp\left\{ -\left[\frac{x - x_0}{\sqrt{2\sigma}} \right]^2 \right\}$$

In this case, the doping profile may represent a field surface implant, or a p-well region. The series resistance per unit length can be represented as

$$r = \frac{\rho_0}{WT_f}$$

where

$$\rho_0 = \frac{1}{q\mu_0 N_A}$$

and T_f is expressed as a function of the number of Gaussian profile standard deviations,

$$T_f = n\sigma$$

Troutman showed that this can be represented as a dimensionless factor n which is a function of the Gaussian profile, or

$$n = \sqrt{\frac{\pi}{2}}\left[1 + erf\left(\frac{\mu}{\sqrt{2\sigma}}\right)\right] + x_1\frac{N_{epi}}{\sigma N_A}$$

In the case of a p^{++} substrate, or a p^{++} buried layer, there is a conductive shunt to the lower 'plate' of the transmission line. Let us define a shunt conductance per unit length of

$$g = \frac{W}{\rho_{epi}(T_{epi} + \lambda)}$$

where the epitaxial resistivity is

$$\rho_{epi} = \frac{1}{q\mu_p N_{epi}}$$

and the effective thickness is determined by the epitaxial flat zone at the end of the surface implant (or p-well), and the beginning of the p^+ substrate (or p^{++} buried layer). In the p^{++} substrate model case, Troutman added an effective depth factor to address the extra effective thickness between the epitaxial flat zone and the heavily doped flat zone of the p^{++} wafer to add an increase in the epitaxial effective resistance.

For the p^{++} wafer, a concentration profile can be assumed to be equal to

$$N_{sub}(x) = \frac{N_{sub}}{2}erfc\left[\frac{x - x_e}{x_0}\right]$$

where we determine the epitaxial depth position from above expression, letting $x_2 = x_e$. From this we can express the effective depth parameter as

$$\lambda = \frac{x_0}{2} erfc^{-1}\left\{\frac{2N_{epi}}{N_{sub}}\right\}$$

Substituting these expressions into the loss factor and characteristic impedance relationships

$$\gamma = \frac{1}{\sqrt{rg}} = T_{epi}\left[\frac{T_f \rho_{epi}\left(1 + \lambda/T_{epi}\right)}{T_{epi}\rho_0}\right]^{1/2}$$

and

$$Z_0 = \sqrt{\frac{r}{g}} = \frac{\rho_{epi}\left(T_{epi} + \lambda\right)}{W\gamma}$$

For the case of the p^{++} buried layer implant, the model can be modified by treating the buried layer as a second series resistance term in parallel with the first series resistance term. In this case, the shunt conductance term would not include the effective depth factor. In the buried layer analysis, the second series resistance term determines the series resistance by integration over the buried layer implant. The buried layer implant would be represented by a two-sided Gaussian profile and the integration of the series resistance would integrate over the conductance. The epitaxial thickness boundary point would coincide with the bounds of the integration of the buried layer implant.

The terminations of the lossy transmission line is a function of the process implants. In the case of the heavily doped buried layer implant, a heavy doped implant can be used to connect the layer to the substrate contact, providing a low resistance shunt between the surface and the buried layer. This provides a low resistance termination of the line for ESD and latchup events.

4.6 SUBSTRATE ABSORPTION, REFLECTION AND TRANSMISSION

Minority-carrier injection into the substrate region influences ESD devices, CMOS latchup and noise. In a substrate region, there are doping concentration gradients vertically and laterally. The substrate region can also have doping concentration variation with transitions between those regions. At the device surface, structures can interfere in the transport of minority carriers across the substrate. With the complexity of integrated electronics, or the structures, to evaluate these analytically in the lateral dimension can be quite difficult. In many problems, we are not interested in the carriers between, but between a first injecting source region at $x = 0$, and a second collecting region at some point $x = L$.

Analytical methods to determine the lateral effects of the substrate are difficult when there are physical regions which are preventing the current flow, or collecting excess minority carriers. For example, *n*-diffusions such as *n*-channel MOSFETs, *n*-wells, and

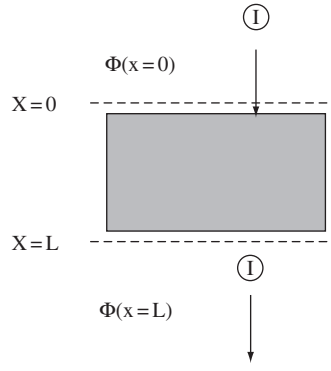

Figure 4.7 Transfer matrix representation of the substrate

n-band triple-well structures can collect minority-carrier electrons in the substrate leading to the collection of the carriers. Additionally, minority-carrier recombination can occur in buried layers, guard rings and other heavily doped p^{++} regions within the physical substrate. Deep trench guard ring structures can also prevent lateral current flow as well as a region of surface recombination.

A method to evaluate the minority carrier transport in the substrate, we can propose a first method treating the region between two points in the substrate. A first method is to regard each physical domain as an independent domain with a defined width and a first and second surface. This modifies the region from a single domain to a stratified medium. From this approach, a matrix approach can be established where the probability that an electron travels from $x = 0$ to $x = L$ is the product of the probabilities of traversing the *n*-domains of the *n*-region stratified medium. In order to evaluate the net probability, the transmission coefficient from a first surface to a second surface is required. Finding the transmission coefficient from the first and second boundary, the net transmission is the product of the transmission coefficients or matrices. In the domains, the carriers are either absorbed, reflected or transmitted. Absorption occurs when the carriers are collected at a metallurgical junction, or recombine. Reflection can occur when the interface is a trench interface. Carriers that are transmitted from a first surface to a second without recombination or collection essentially 'escape', leading to transmission.

This method determines the minority carrier current from a first point to a second point and can accurately address surface structures which are typically ignored in the substrate analysis methods. This method does not determine the voltage and current potentials as is achieved in the substrate transmission line model method.

In semiconductor wafers, there are uniform profiles with low doping concentration. There are also wafers with substrate wafers with high doping concentration, and thin-film low-doped epitaxial films.

4.7 SUBSTRATE ELECTRICAL AND THERMAL DISCRETIZATION

For ESD simulation, the modeling of the substrate is important for understanding thermal and electrical coupling effects [14]. For the electrical properties, we can state

$$\nabla \bullet (\varepsilon \nabla \psi) = -q(p - n + N_{\mathrm{D}} + N_{\mathrm{A}})$$

$$\frac{\partial n}{\partial t} = +\frac{1}{q}\nabla \bullet J_n + G(n,p) - R(n,p)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \bullet J_p + G(n,p) - R(n,p)$$

and the thermal relationship can be shown as

$$c\frac{\partial T}{\partial t} + \nabla \bullet J_Q = H$$

where H is the Joule heating, the Thomson heating and recombination–generation heating drive terms. Near the surface of a device, these terms are significant, but in the substrate region, the following terms can be regarded as small compared with the terms on the left-hand side of the heat equation

$$E = \frac{{J_n}^2}{\sigma_n} + \frac{{J_p}^2}{\sigma_p}$$

$$E = T\{\nabla P_n \bullet J_n - \nabla P_p \bullet J_p\}$$

$$E = qR[\phi_p - \phi_n + T(P_p + P_n)]$$

then we can approximate the substrate interaction as a homogeneous differential equation with variable coefficients

$$c\frac{\partial T}{\partial t} + \nabla \bullet J_Q = 0$$

This development assumes that the Joule heating, and recombination energy is small compared with the heat flux and temperature gradient. From this relationship, the divergence of the heat flux density can be expressed as

$$\nabla \bullet J_Q = -c\frac{\partial T}{\partial t}$$

For an infinitesimal volume, it is known from the divergence theorem, that the divergence of the heat flux is equal to the heat flux through the surface. Consider a control volume surrounding a nodal point i, then

$$\iiint_{\Omega_i} \nabla \bullet J \mathrm{d}\Omega = \iint_{S_i} J \mathrm{d}S = \iiint_{\Omega_i} -\left[c\frac{\partial T}{\partial t}\right] \mathrm{d}\Omega$$

Hence

$$\iint_{S_i} J \mathrm{d}S = \iiint_{\Omega_i} \left[-c \frac{\partial T}{\partial t} \right] \mathrm{d}\Omega$$

This expression can be put into a discrete formulation between node i and all nodes j which are nearest neighbors. Then we can put in the form

$$\iint_{S_i} J \mathrm{d}S = \sum_j J_{ij} S_{ij} = \sum_j J_{ij} w_{ij} d_{ij} = \left\langle \left[-c \frac{\partial T}{\partial t} \right]_i \right\rangle \Omega_i$$

where the surface between node i and node j is represented as dimensions w, and d.

From this form, we can also state

$$\nabla \bullet J = \frac{1}{\Omega_i} \sum_j J_{ij} w_{ij} d_{ij}$$

From the heat conduction equation, we can apply the discretization over a volume element as

$$\Omega_i c_i \frac{\partial T_i}{\partial t} + \Omega_i \langle \nabla \bullet J_Q \rangle_i = 0$$

Then, substituting in the expression for the divergence of the thermal flux

$$\Omega_i c_i \frac{\partial T_i}{\partial t} + \Omega_i \left\{ \frac{1}{\Omega_i} \sum_j J_{ij} w_{ij} d_{ij} \right\} = 0$$

Expressing the incremental heat flux as the local thermal conductivity of the ith element, and the temperature differential between the temperature at node i to node j divided by the spacing between the two nodal points,

$$J_{ij} = -k_i \frac{T_i - T_j}{h_{ij}}$$

Substituting in the discrete form of the heat flux

$$\Omega_i c_i \frac{\partial T_i}{\partial t} + \Omega_i \left\{ \frac{1}{\Omega_i} \sum_j k_i \frac{T_i - T_j}{h_{ij}} w_{ij} d_{ij} \right\} = 0$$

Regrouping the terms

$$\{\Omega_i c_i\}\frac{\partial T_i}{\partial t} + \left\{\sum_j k_i \frac{w_{ij} d_{ij}}{h_{ij}}(T_i - T_j)\right\} = 0$$

From the above expression, we can define the thermal capacitance for a volume element as

$$C_i = \Omega_i c_i$$

and thermal conductance for the volume element i from the ith element to the jth element

$$G_{ij} = k_i \frac{w_{ij} d_{ij}}{h_{ij}}$$

$$C_i \frac{\partial T_i}{\partial t} + \left\{\sum_j G_{ij}(T_i - T_j)\right\} = 0$$

This expression for a nodal point i in a discretized mesh remains valid assuming it can be used to define the thermal capacitance and thermal conductances at a given point. In a three-dimensional rectangular mesh, the electrical analog is six conductance elements from the node i to all nodes j in each direction, and a capacitor element whose electrode is from node i to the thermal sink plane.

In this form, we can relate the substrate formulation for the thermal representation and its associated electrical representation. From the constitutive relationships we have shown

$$J_n = -q\mu_n nE + qD_n \nabla n + q_n {D_n}^T \nabla T$$

$$J_p = -q\mu_p pE - qD_p \nabla p + q_p Dp^T \nabla T$$

Normalizing the constitutive relationship for electrons, and substituting in the electrical field as the gradient of the potential

$$\frac{J_n}{q\mu_n n} = -\nabla V + \frac{qD_n}{q\mu_n n}\nabla n + \frac{q_n Dn}{q\mu_n n}\nabla T$$

Letting

$$\frac{J_n}{q\mu_n n} = -\nabla V + \frac{D_n}{\mu_n}\frac{\nabla_n}{n} + \frac{D_n}{\mu_n}\nabla T$$

or

$$\frac{J_n}{q\mu_n n V_{\text{th}}} = -\frac{\nabla V}{V_{\text{th}}} + \frac{\nabla n}{n} + \nabla T$$

From this form, we can make the assumption that, if over an incremental distance the normalized electrical potential change is significantly greater than the normalized change in the carrier density or the thermal gradient, then the system reduces to a drift-dominated system, where the diffusion current and temperature gradient drive terms are assumed negligible.

The carrier transport expressions can be assumed to be drift dominated, where

$$J_n = -q\mu_n n \nabla V$$

$$J_p = -q\mu_p p \nabla V$$

From the current continuity expression,

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla \bullet J_n = G - R$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla \bullet J_p = G - R$$

We can show for an *n*-type substrate, where ρ is the resistivity

$$q\left\{\frac{\partial p}{\partial t} - \frac{\partial n}{\partial t}\right\} = -\nabla \bullet \left\{J_n + J_p\right\} = \frac{1}{\rho}\{\nabla \bullet E\}$$

By differentiation of Gauss' law for semiconductors with respect to time

$$\frac{\partial}{\partial t}\nabla \bullet (\varepsilon \nabla \psi) = \frac{\partial}{\partial t}\{-q(p - n + N_{\text{D}} + N_{\text{A}})\} = -q\left\{\frac{\partial p}{\partial t} - \frac{\partial n}{\partial t}\right\}$$

From the two above equations, the time and spatial variation of the electric field can be shown to be equal to

$$\varepsilon \frac{\partial}{\partial t}\{\nabla \bullet E\} + \frac{1}{\rho}\{\nabla \bullet E\} = 0$$

From this form, it is evident that the temporal response of the substrate is associated with the relationship of the characteristic time scale and the charge relaxation time. Putting the expression in a normalized time formulation it is evident that

$$\frac{\partial}{\partial(t/\varepsilon\rho)}\{\nabla \bullet E\} + \{\nabla \bullet E\} = 0$$

The charge relaxation time is a microscopic form of the effective *RC* time of the medium. The discretization of this expression can be formed where

$$\nabla \bullet E = \frac{1}{\Omega_i} \sum_j E_{ij} w_{ij} d_{ij}$$

such that

$$\varepsilon \frac{\partial}{\partial t} \left\{ \frac{1}{\Omega_i} \sum_j E_{ij} w_{ij} d_{ij} \right\} + \frac{1}{\rho} \left\{ \frac{1}{\Omega_i} \sum_j E_{ij} w_{ij} d_{ij} \right\} = 0$$

Substituting for the potential difference for a incremental surface between

$$E_{ij} = \frac{V_i - V_j}{h_{ij}}$$

$$\left\{ \frac{1}{\Omega_i} \sum_j \varepsilon w_{ij} d_{ij} \frac{\partial}{\partial t} \left\{ \frac{V_i - V_j}{h_{ij}} \right\} \right\} + \left\{ \frac{1}{\Omega_i} \sum_j \sigma w_{ij} d_{ij} \left\{ \frac{V_i - V_j}{h_{ij}} \right\} \right\} = 0$$

Then, we can express the system as

$$\sum_j G_{ij} \left(V_i - V_j \right) + C_{ij} \left(\frac{\partial V_i}{\partial t} - \frac{\partial V_j}{\partial t} \right) = 0$$

where we define the electrical conductance and capacitance terms as

$$G_{ij} = \sigma \frac{w_{ij} d_{ij}}{h_{ij}}$$

$$C_{ij} = \varepsilon \frac{w_{ij} d_{ij}}{h_{ij}}$$

In this form, the conductance and capacitance terms are in parallel for each of the six faces of the volumetric region from nodal point *i* to all adjacent nodes *j*. Hence, the electrical model of the substrate is associated with the effective *RC* time of the substrate. The dynamic of the substrate is a function of the characteristic time of interest relative to the effective *RC* time of the substrate. If the conductivity of the substrate is large, the carriers can screen out the electrical field allowing for the medium to establish equilibrium. For highly resistive substrates, the electrical fields require a longer time to damp the transient electrical fields.

4.8 SUBSTRATE EFFECTS: ELECTRICAL TRANSFER RESISTANCE

For electrostatic discharge phenomenon, the modeling of the substrate is important for understanding thermal and electrical coupling effects. Troutman applied the concept of

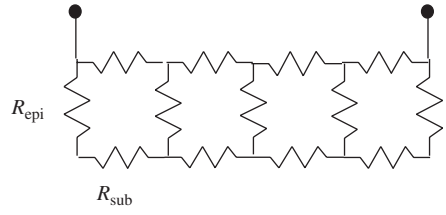

Figure 4.8 Transfer resistance representation of substrates as a ladder network

transfer resistance to the substrate to address latchup-related coupling phenomenon [6]. In two dimensions, the network can be represented as a ladder network (Figure 4.8).

Figure 4.9 addresses the case of three dimensions and multiple elements in a common substrate. When there are multiple elements contained within a common substrate, the substrate potential for a given element is dependent on the voltage drops or current in the system [15].

Given n elements contained in a common substrate, we can define a general system with a relationship to n currents as

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ \cdots \\ Vn \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & \cdots & R_{1n} \\ R_{21} & R_{22} & & & R_{2n} \\ \cdots & & \cdots & & \\ \cdots & & & \cdots & \\ R_{n1} & R_{n2} & \cdots & \cdots & R_{nn} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdots \\ \cdots \\ I_n \end{bmatrix}$$

Representing the resistance expression as a second-order tensor with and the voltage and current as a first-order vector

$$V_i = R_{ij} I_j$$

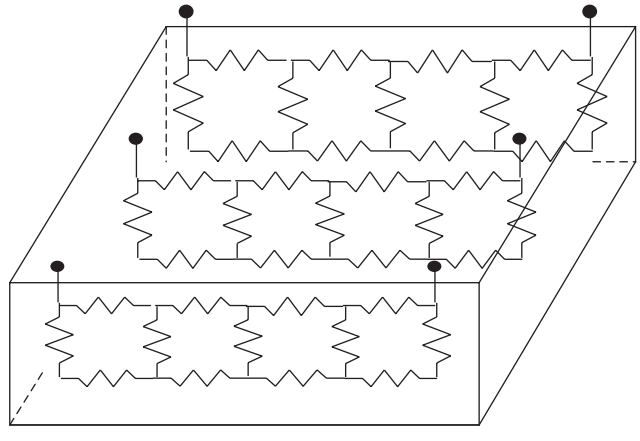

Figure 4.9 Three-dimensional resistive substrate matrix with a plurality of inputs

or

$$V_i = \sum_{j=1}^{j=n} R_{ij} I_j = R_{i1} I_1 + R_{i2} I_2 + R_{i3} I_3 + \ldots + R_{in} I_n$$

In this formulation, the resistance terms of the main diagonals are 'self-resistance' terms and the off-diagonal elements are 'mutual resistance' terms or transfer resistance terms. In this form, the influence on the ith element is a function of the current injected into the substrate from the other elements.

In the case where there is a local substrate contact for every element, and weak coupling, then we can assume that

$$R_{ij} = 0 \quad \forall i \neq j$$

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ \cdots \\ V_n \end{bmatrix} \simeq \begin{bmatrix} R_{11} & 0 & \cdots & \cdots & 0 \\ 0 & R_{22} & & & \\ \cdots & & \cdots & & \\ \cdots & & & \cdots & \\ 0 & \cdots & \cdots & \cdots & R_{nn} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdots \\ \cdots \\ I_n \end{bmatrix}$$

In this case the individual elements can be treated as independent elements, with no interaction through the substrate.

In the case that there are n elements, but only m elements are connected to the common substrate contact, and all n–m elements have independent substrate contacts, then the form of the matrix can be described as

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ V_m \\ \cdots \\ V_{n-1} \\ V_n \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} & 0 & 0 & 0 \\ R_{21} & R_{22} & & R_{2m} & 0 & 0 & 0 \\ \cdots & & \cdots & & 0 & 0 & 0 \\ R_{m1} & R_{m2} & & R_{mm} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & R_{n-1n-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & R_{nn} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdots \\ \\ \\ I_{n-1} \\ I_n \end{bmatrix}$$

The second-order resistance tensor main diagonal terms can be estimated from the physical resistance of the device to a common plane in the substrate region. This can be achieved through device simulation or analytically. This resistance may be comprised of a series of resistance elements in which the injected current flows. The off-diagonal terms of the second-order resistance tensor can be obtained through simulation or analytically. These are associated with the influence of the substrate potential under the ith element as a result of injected current at the jth element

In the solution by Li *et al.* [15], the transfer resistance was expressed in the form of an $m \times n \times p$ matrix of transfer resistances, where there are m injection current sources, n substrate contacts, and p locations of interest. In this form, it can be expressed as

$$R_{ji}^k = V_{ji}^k / I_i$$

where i is the index for the m injection sources, j is the index of the p locations of interest, and k is for the n substrate contacts. In this form, the transfer matrix terms can be solved for the m injection locations, and n substrate contacts. This can be applied to multiple devices, or multi-finger elements.

4.9 SUBSTRATE EFFECTS: THERMAL TRANSFER RESISTANCE

To understand of the thermal temperature in a multiple element system common to a given substrate, we can analogously derive the following relationship to understand the thermal coupling for the thermal electrical model. When there are multiple elements contained within a common substrate, the substrate local temperature for a given element is dependent on the temperature variation in the system from surrounding elements. The temperature is equal to the product of power and thermal impedance

$$T = P\theta_{\mathrm{TH}}$$

Hence, given n elements contained in a common substrate, we can define a general system with a relationship to n currents as

$$\begin{bmatrix} T_1 \\ T_2 \\ \cdots \\ \cdots \\ T_n \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \cdots & \theta_{1n} \\ \theta_{21} & \theta_{22} & & & \theta_{2n} \\ \cdots & & \cdots & & \cdots \\ \cdots & & & \cdots & \cdots \\ \theta_{n1} & \theta_{n2} & \cdots & \cdots & \theta_{nn} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \cdots \\ \cdots \\ P_n \end{bmatrix}$$

Representing the thermal resistance expression as a second-order tensor and the temperature and power as first-order vectors

$$T_i = \theta_{ij} P_j$$

or

$$T_i = \sum_{j=1}^{j=n} \theta_{ij} P_j = \theta_{i1} P_1 + \theta_{i2} P_2 + \theta_{i3} P_3 + \ldots + \theta_{in} P_n$$

In this formulation, the thermal resistance terms of the main diagonals are 'self-thermal resistance' terms and the off-diagonal elements are 'mutual thermal resistance' terms or transfer thermal resistance terms. In this form, the influence on the ith element is a function of the power injected into the substrate from the other elements.

In the case where there is a local substrate thermal contact for every element, and weak coupling, then we can assume that

$$\theta_{ij} = 0 \quad \forall i \neq j$$

$$\begin{bmatrix} T_1 \\ T_2 \\ \cdots \\ \cdots \\ T_n \end{bmatrix} \simeq \begin{bmatrix} \theta_{11} & 0 & \cdots & \cdots & 0 \\ 0 & \theta_{22} & & & \\ \cdots & & \cdots & & \\ \cdots & & & \cdots & \\ 0 & \cdots & \cdots & \cdots & \theta_{nn} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \cdots \\ \cdots \\ P_n \end{bmatrix}$$

In this case the individual elements can be treated as independent elements, with no interaction through the substrate.

In the case that there are n elements, but only m elements are connected to the common thermal substrate contact, and all n–m elements have independent thermal substrate contacts, then the form of the matrix can be described as

$$\begin{bmatrix} T_1 \\ T_2 \\ \cdots \\ T_m \\ \cdots \\ T_{n-1} \\ T_n \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1m} & 0 & 0 & 0 \\ \theta_{21} & \theta_{22} & & \theta_{2m} & 0 & 0 & 0 \\ \cdots & & \cdots & & 0 & 0 & 0 \\ \theta_{m1} & \theta_{m2} & & \theta_{mm} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{n-1n-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{nn} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \cdots \\ \\ \\ P_{n-1} \\ P_n \end{bmatrix}$$

The second-order thermal resistance tensor main diagonal terms can be estimated from the physical thermal resistance of the device to a common plane in the substrate region. This can be achieved through device simulation or analytically. This thermal resistance may be comprised of a series of thermal resistance elements in which the injected power flows. The off-diagonal terms of the second-order thermal resistance tensor can be obtained through simulation or analytically. These are associated with the influence of the thermal substrate potential under the ith element as a result of injected power at the jth element.

Analogous to the method of Li [15], the thermal transfer resistance can be expressed in the form of an $m \times n \times p$ matrix of transfer resistances, where there are m injection power sources, n thermal sinks, and p locations of interest. In this form, it can be expressed as

$$\theta_{ji}^{k} = T_{ji}^{k} / P_i$$

where i is the index for the m thermal sources, j is the index of the p locations of interest, and k is that for the n substrate thermal contacts

4.10 SUBSTRATE THERMAL RESISTANCE MODELS

4.10.1 Variable Cross-section Model

Substrate resistance models are important to evaluate the electrical resistance, and thermal resistance for evaluation of the power to failure and peak temperature [16–19]. When a semiconductor device discharges current into a semi-infinite region, the current density as a function of angle is nonuniform when the substrate contact

region is in the z-direction. The current density decreases as a function of rotational angle, leading to a region where the current density is negligible. In this analysis, the boundary conditions are simplified to allow a simple solution. For the purposes of modeling, a conical section can be defined which assumes an adiabatic virtual boundary which does not allow the flow of current outside some defined cross section $A(z)$ which is variable in the z-direction [18,19]. This method avoids Green's functions, and thermal mixed boundary conditions.

Additionally, near the surface of a semi-infinite region, isolation regions exist which prevent the lateral current transport whose cross-section regions are variable in the z-direction. Semiconductor structures are defined by LOCOS isolation, shallow trench isolation (STI), deep trench (DT) and V-groove structures. The region in which the current is flowing can be of increasing area or decreasing area, dependent on the slope of the region sidewalls. For the electrical current, the current does not flow through the physical boundaries. For thermal effects, if the heat flux is assumed negligible at short time scales, it can be assumed the perimeter is adiabatic and all heat flux does not flow through the insulating regions.

Generalizing the expression for thermal resistance, under the above assumptions, we can define the thermal resistance as a series or integral of thermal resistances with a variable cross-section,

$$R_{\mathrm{TH}} = \int \frac{\mathrm{d}z}{\kappa(z)A(z)}$$

Assuming a rectangular parallel surface of length l and width w, a three-dimensional trapezoidal region can be defined such that the area increases in the z-direction as a fixed angle in both the directions, where the length and width are increasing with the same defined angle. The length and width follow the form

$$w(z) = w_0 + 2z\tan\theta$$

$$l(z) = l_0 + 2z\tan\theta$$

where the form of the thermal resistor is

$$R_{\mathrm{TH}} = \int_0^L \frac{\mathrm{d}z}{\kappa(z)(w_{\mathrm{o}} + 2z\tan\theta)(l_{\mathrm{o}} + 2z\tan\theta)}$$

The thermal resistor expression above assumes the area is increasing with depth. This development can be used in a region which is decreasing in area with the dimension z as well (e.g. silicon region in a V-groove insulator or sloped sidewall).

The above expression can be expressed as

$$R_{\mathrm{TH}} = \int_0^L \frac{\mathrm{d}z}{4\kappa(z)\tan^2\theta(z+\alpha)(z+\beta)}$$

where

$$\alpha = \frac{w_0}{2\tan\theta}$$

and

$$\beta = \frac{l_0}{2\tan\theta}$$

From the integration method of partial fractions, the undetermined coefficients are obtained, and the integral can be put in the following form

$$R_{\mathrm{TH}} = \int_0^L \frac{\mathrm{d}z}{4\kappa(z)\tan^2\theta}\left\{\frac{1}{\beta-\alpha}\right\}\left\{\frac{1}{z+\alpha}-\frac{1}{z+\beta}\right\}$$

In this form, it is clear that the integrals can be represented as natural logarithms. To simplify the form, we will assume the thermal conductivity is not a function of depth. Solving the integrals and applying the identity for the subtraction of natural logarithms

$$R_{\mathrm{TH}} = \frac{1}{4\kappa\tan^2\theta}\left[\frac{1}{\beta-\alpha}\ln\left\{\frac{z+\alpha}{z+\beta}\right\}\right]_{0|}^{L} = \frac{1}{4\kappa\tan^2\theta}\frac{1}{\beta-\alpha}\left\{\ln\left[\frac{L+\alpha}{L+\beta}\right]-\ln\left[\frac{\alpha}{\beta}\right]\right\}$$

Applying the identity for the subtraction of natural logarithms and substituting in for the variables, the general expression can be expressed as

$$R_{\mathrm{TH}} = \frac{1}{2\kappa\tan\theta}\left(\frac{1}{l_0-w_0}\right)\ln\left[\frac{(2L+w_0/\tan\theta)}{(2L+l_0/\tan\theta)}\frac{(l_0/\tan\theta)}{(w_0/\tan\theta)}\right]$$

This expression can be simplified in a few cases. For the first case assume that the length is significantly greater than the width expression
Case I: $2L \gg w_0/\tan\theta$

$$R_{\mathrm{TH}} = \frac{1}{2\kappa\tan\theta}\left(\frac{1}{l_0-w_0}\right)\ln\left[\frac{l_0}{w_0}\xi_l\right]$$

where

$$\xi_l = \frac{1}{1+\left(\frac{l_0}{2L\tan\theta}\right)}$$

Case II: $2L \gg w_0/\tan\theta$ and $2L \gg l_0/\tan\theta$, the expression simplifies to the form

$$R_{\mathrm{TH}} = \frac{1}{2\kappa\tan\theta}\left(\frac{1}{l_0-w_0}\right)\ln\left[\frac{l_0}{w_0}\right]$$

The integral expression can be approximated as a Taylor expansion

$$\ln x \simeq \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \cdots$$

The expression can be simplified as

$$R_{\mathrm{TH}} = \frac{1}{2\kappa \tan\theta}\left(\frac{1}{l_0 - w_0}\right)\ln\left[\frac{l_0}{w_0}\right] \simeq \frac{1}{2\kappa \tan\theta}\left(\frac{1}{l_0 - w_0}\right)\left[\frac{l_0 - w_0}{l_0} + \frac{1}{2}\left(\frac{l_0 - w_0}{l_0}\right)^2\right]$$

$$R_{\mathrm{TH}} = \frac{1}{2\kappa \tan\theta}\left(\frac{1}{l_0 - w_0}\right)\ln\left[\frac{l_0}{w_0}\right] \simeq \frac{1}{2\kappa l_0 \tan\theta}$$

This shows that the result is associated with a characteristic dimension of the device. In another development using a Green's function approach, the simplification of the expression reduced to the form

$$R_{\mathrm{TH}} = \frac{1}{4\kappa\sqrt{l_0 w_0}}$$

From the Green's function approach of Joy and Schlig [1], the characteristic length is the arithmetic mean of the width and length.

4.10.2 Variable Elliptical Cross-section Model

Evaluation of the thermal resistance for an elliptical region with variable cross-section as a function of depth with insulating sidewalls can be obtained in the same fashion as the rectangular structure. The equation of the ellipse in the x- and y-directions where z is a parameter of the minor and major axis, is

$$\frac{x^2}{a^2(z)} + \frac{y^2}{b^2(z)} = 1$$

The area of the ellipse is calculated by integration with respect to x

$$A(z) = 2\int_{x=-a(z)}^{x=a(z)} |y(x,z)|\mathrm{d}x = 4\int_{x=0}^{x=a(z)} b(z)\sqrt{1 - \frac{x^2}{a(z)^2}}\,\mathrm{d}x$$

$$A(z) = 4\frac{b(z)}{a(z)}\int_{x=0}^{x=a(z)} \sqrt{a^2(z) - x^2}\,\mathrm{d}x = \frac{4b(z)}{a(z)}\left[\frac{x}{2}\sqrt{a^2 - x^2} + \frac{a^2(z)}{2}\sin^{-1}\left(\frac{x}{a}\right)\right]\Bigg|_0^a = \pi a(z) b(z)$$

Generalizing the expression for thermal resistance, under the above assumptions, we can define the thermal resistance as a series or integral of thermal resistances with a variable cross-section

$$R_{\mathrm{TH}} = \int \frac{\mathrm{d}z}{\kappa(z)A(z)} = \int_0^L \frac{\mathrm{d}z}{\pi\kappa(z)\{a(z)b(z)\}}$$

A three-dimensional trapezoidal region can be defined such that the area increases in the z-direction as a fixed angle in both the directions

$$a(z) = a_0 + 2z\tan\theta$$

$$b(z) = b_0 + 2z\tan\theta$$

The above expression can be expressed as

$$R_{\mathrm{TH}} = \int_0^L \frac{\mathrm{d}z}{\pi\kappa(z)\tan^2\theta(z+\alpha)(z+\beta)}$$

where

$$\alpha = \frac{a_0}{\tan\theta}$$

and

$$\beta = \frac{b_0}{\tan\theta}$$

From the integration method of partial fractions, the undetermined coefficients are obtained, and the integral can be put in the following form

$$R_{\mathrm{TH}} = \int_0^L \frac{\mathrm{d}z}{\pi\kappa(z)\tan^2\theta}\left\{\frac{1}{\beta-\alpha}\right\}\left\{\frac{1}{z+\alpha} - \frac{1}{z+\beta}\right\}$$

In this form, it is clear that the integrals can be represented as natural logarithms. To simplify the form, we will assume the thermal conductivity is not a function of depth. Solving the integrals and applying the identity for the subtraction of natural logarithms

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan^2\theta}\left[\frac{1}{\beta-\alpha}\ln\left\{\frac{z+\alpha}{z+\beta}\right\}\right]_{0|}^{L} = \frac{1}{4\kappa\tan^2\theta}\frac{1}{\beta-\alpha}\left\{\ln\left[\frac{L+\alpha}{L+\beta}\right] - \ln\left[\frac{\alpha}{\beta}\right]\right\}$$

Applying the identity for the subtraction of natural logarithms and substituting in for the variables, the general expression can be expressed as

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan\theta}\left(\frac{1}{b_0-a_0}\right)\ln\left[\frac{(L\tan\theta+a_0)}{(L\tan\theta+b_0)}\frac{(b_0)}{(a_0)}\right]$$

This expression can be simplified in a few cases. For the first case assume that the length is significantly greater than the width expression
Case I: $2L \gg a_0/\tan\theta$

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan\theta}\left(\frac{1}{b_0 - a_0}\right)\ln\left[\frac{b_0}{a_0}\xi_b\right]$$

where

$$\xi_b = \frac{1}{1 + \left(\frac{b_0}{L\tan\theta}\right)}$$

Case II: $2L \gg a_0/\tan\theta$ and $2L \gg b_0/\tan\theta$, the expression simplifies to the form

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan\theta}\left(\frac{1}{b_0 - a_0}\right)\ln\left[\frac{b_0}{a_0}\right]$$

The integral expression can be approximated as a Taylor expansion

$$\ln x \simeq \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \cdots$$

The expression can be simplified as

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan\theta}\left(\frac{1}{b_0 - a_0}\right)\ln\left[\frac{b_0}{a_0}\right] \simeq \frac{1}{\pi\kappa\tan\theta}\left(\frac{1}{b_0 - a_0}\right)\left[\frac{b_0 - a_0}{b_0} + \frac{1}{2}\left(\frac{b_0 - a_0}{b_0}\right)^2\right]$$

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan\theta}\left(\frac{1}{b_0 - a_0}\right)\ln\left[\frac{b_0}{a_0}\right] \simeq \frac{1}{\pi\kappa b_0\tan\theta}$$

For the case of a 45° angle

$$R_{\mathrm{TH}} \simeq \frac{1}{\pi\kappa b}$$

For the case of a conical region with insulating sidewalls (Figure 4.10)

$$R_{\mathrm{TH}} = \int \frac{\mathrm{d}z}{\kappa(z)A(z)} = \int_0^L \frac{\mathrm{d}z}{\kappa(z)\{\pi r^2(z)\}} = \int_0^L \frac{\mathrm{d}z}{\kappa(z)\left\{\pi(r_0 + z\tan\theta)^2\right\}}$$

$$R_{\mathrm{TH}} = \frac{1}{\pi\kappa\tan^2\theta}\int_0^L \frac{\mathrm{d}z}{(z + r/\tan\theta)^2} = \frac{1}{\pi\kappa\tan^2\theta}\left\{\left(\frac{L}{L + r_0/\tan\theta}\right)\left(\frac{1}{r_0/\tan\theta}\right)\right\}$$

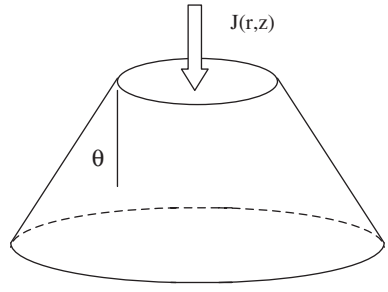

Figure 4.10 Thermal resistance substrate model with a conical insulating boundary condition

When the substrate length L is significantly larger than the radius

$$R_{\mathrm{TH}} \simeq \frac{1}{\pi \kappa r_0 \tan \theta}$$

where for the angle of 45 °

$$R_{\mathrm{TH}} \simeq \frac{1}{\pi \kappa r_0}$$

Note that this solution is consistent with the elliptical solution. From the elliptical solution and the Taylor expansion of the natural logarithm solution, in the limit that the minor and major axis are equal, the result, where a approaches b, is

$$R_{\mathrm{TH}} = \frac{1}{\pi \kappa}\left(\frac{1}{b_0 - a_0}\right) \ln\left[\frac{b_0}{a_0}\right] \simeq \frac{1}{\pi \kappa}\frac{1}{a_0}\left(\frac{1}{b/a^{-1}}\right)\left[\frac{b/a - 1}{b/a} + \ldots\right] = \frac{1}{\pi \kappa b}$$

From this derivation, we have shown that, for the case of an elliptical region or circular region whose current spreading can be treated as a virtual where a minimal current flow passes, or a real insulator boundary which does not pass current, a solution exists for both the electrical and thermal problems. In all cases, when the length of the elliptical or circular region is significantly larger than the size of the plane dimensions, the thermal (or electrical) resistance is associated with a characteristic length of the plane dimension.

4.10.3 Back-surface Substrate Lumped Analytical Model

A lumped analytical model for substrate resistance associated with a back-surface contact is given by Schutz, Selberherr, and Potzl, using a substrate current spreading angle θ, a wafer thickness t, width parameter w, and silicon resistivity ρ_s[19]. This substrate resistance term can be expressed as

$$R_{\mathrm{sub}} = \frac{\rho_s}{2 \tan \theta} \ln\left[\frac{w(t)}{w}\right]$$

where the width forms a conical region to the back of the semiconductor chip. In this fashion, we can express the width of the region, as a function of the spreading angle of current

$$w(t) = w + 2t\tan\theta$$

where

$$R_{\text{sub}} = \frac{\rho_{\text{s}}}{2\tan\theta}\ln\left[\frac{w + 2t\tan\theta}{w}\right]$$

The model is for evaluation of the electrical substrate resistance for a back-surface contact. This model can be adapted to models of the device on the surface to evaluate the substrate resistance effect on the device when a back-surface contact is present.

4.11 HEAVILY DOPED SUBSTRATES

Heavily doped substrate wafers have a significant effect on the ESD robustness of semiconductor components and ESD networks. Heavily doped wafers have significant advantages for semiconductor chips. They have short recombination times due to the Auger recombination process. As a result, they provide good resilience to ionized charge events such as α-particles, cosmic rays and heavy ion events. Heavily doped substrates offer low soft error rate (SER), and low sensitivity to CMOS latchup. Heavily doped substrates also lower the requirement and sensitivity of substrate contact spacing and design. Heavily doped substrate wafers also had a significant effect on ESD design and results. The heavily doped substrate provided a low resistance shunt to the substrate which was an advantage for some ESD designs and disadvantage for others [7–13].

The p^{++} wafer can be represented as a concentration profile of the form of the complementary error function

$$N_{\text{sub}}(x) = N_{\text{epi}} + \frac{N_{\text{sub}}}{2}erfc\left[\frac{x - x_{\text{e}}}{x_0}\right]$$

In the case of the MOSFET device, the heavily doped substrate led to an increase in the MOSFET snapback voltage. In order for MOSFET snapback to occur, the source-to-substrate voltage must be established, and it must be greater than or equal to the product of the substrate current times the substrate resistance. From the current equation condition for latchup

$$V_{\text{be}} \geqslant I_{\text{sx}}R_{\text{sub}}$$

As the substrate resistance is decreased, the MOSFET must achieve a higher substrate current to undergo MOSFET snapback. Hence, from the MOSFET current relationship

$$V_{\text{be}} \geqslant [(M-1)(I_{\text{DS}} + I_{\text{C}}) - I_{\text{B}}]R_{\text{sub}}$$

As a result the first term in the bracket must increase significantly as the substrate resistance is decreased. Additionally, in a multi-finger MOSFET structure, because only some of the electrical nodes are coupled, different MOSFET fingers do not turn on all at the same voltage. It is an advantage in this case to have the thermal instability voltage Vt_2 exceed the electrical instability voltage Vt_1. When MOSFET snapback, Vt_1 exceeds Vt_2, all MOSFET fingers do not turn on. Hence, the ESD robustness of a multi-finger MOSFET decreases. As a result of this effect, the ESD robustness of a MOSFET in a p^+ substrate is significantly lower than in a *p*-substrate wafer.

In the case of a diode structure, the ESD robustness of diode structures significantly improves with a p^{++} substrate wafer. Using an *n*-well-to-substrate diode element, the p^{++} substrate region provides a low diode series resistance to the bulk substrate region. The *n*-well-to-substrate metallurgical junctions are designed such that the *n*-well regions abut the transition doping concentration between the epitaxy doping concentration flat zone, and the heavily doped substrate flat zone. In these structures, no substrate contacts are needed local to the *n*-well diode. Human body measurements of *n*-well diodes with STI-defined n^+ contacts in a p^+ substrate were first measured by Voldman and Hargrove. ESD human body model results exceeding -10 kV were achieved with a 100-μm-long diode. Brown and Voldman demonstrated that the HBM ESD robustness of the *n*-well diodes reduced to -5 kV on a *p*-substrate wafer. Hence, the p^{++} substrate provides at least a two-fold improvement in the ESD robustness of *n*-well diode structures compared with a *p*-substrate wafer. A second advantage of the p^{++} heavily doped wafer is that the vertical bipolar transistor will be less sensitive to the Kirk effect. In the vertical *pnp*, as the base current flows from the p^+ source/drain implant to the p^{++} substrate, the *n*-well-to-substrate metallurgical junction will be less sensitive to base push-out effects, preventing the reduction in the unity current gain cutoff frequency. Additionally, for the case of the vertical *pnp*, the reverse current transport term α_r is high as a result of the high substrate doping concentration relative to the emitter. This provides good reverse injection when the substrate is at a ground reference potential, and a negative pulse is applied to the input node. This effect was observed from the failure mechanisms in STI-bound double-diode networks. This subtle effect was observed from the interconnect damage patterns at -10 kV in diode structures.

There are many subtle advantages of heavily doped substrates which are advantageous for ESD and latchup. For example, an additional advantage of p^{++} heavily doped substrates is the improvement of guard ring efficiency. Guard rings provide the ability to collect minority-carrier electrons; *n*-well diodes, or other STI-bound diffusions will inject electrons in the epitaxial region. By placement of a *n*-well ring around these structures, minority carrier electrons either enter the p^{++} substrate region or are collected by the *n*-well ring structure. The minority carrier electrons which enter the p^{++} substrate have a low probability of escape outside of the *n*-well ring structure because of the high recombination rate in the p^+ substrate.

4.12 LOW-DOPED SUBSTRATES

Low-doped substrate wafers have a significant effect on the ESD robustness of semiconductor components and ESD networks. Low-doped p^- wafers have significant advantages for semiconductor chips. Low doped p^- wafers provide low substrate

capacitance, good noise isolation, and reduced cost. This is an advantage for semiconductor components from bipolar transistors to MOSFETs. In the case of bipolar transistors, the lowering of the collector-to-substrate capacitance leads to a significant increase in the unity power gain cutoff frequency. Noise isolation is improved between analog and digital circuitry. Additionally, for RF applications, high quality factor (Q) passive inductors can be constructed on low-doped substrates.

For the case of the p-well near the device surface, with a background doping of N_{sub}, we can define a concentration for the surface region as

$$N_{\text{A}}(x) = N_{\text{sub}} + N_{\text{A}} \exp\left\{ -\left[\frac{x - x_0}{\sqrt{2\sigma}}\right]^2 \right\}$$

In the case of the MOSFET device, the lightly doped substrate lead to a lowering of the MOSFET snapback voltage. In order for MOSFET snapback to occur, the source-to-substrate voltage must be established and it must be greater than or equal to the product of the substrate current times the substrate resistance. From the current equation condition for MOSFETs

$$V_{\text{be}} \geqslant I_{\text{sx}} R_{\text{sub}}$$

As the substrate resistance is increased, the MOSFET can have a lower substrate current to trigger MOSFET snapback. Hence, from the MOSFET current relationship

$$V_{\text{be}} \geqslant [(M-1)(I_{\text{DS}} + I_{\text{C}}) - I_{\text{B}}] R_{\text{sub}}$$

As a result the first term in the bracket can be reduced significantly as the substrate resistance is increased. Additionally, in a multi-finger MOSFET structure, ESD robustness is improved because the MOSFET first trigger voltage V_{t1} becomes lower than the MOSFET second trigger voltage V_{t2} value. As a result of this effect, the ESD robustness of a MOSFET in a p^- substrate is significantly higher.

As technologies are scaled to lower voltages, the scaling of the substrate doping concentration will lead to lower MOSFET snapback voltages, which is an advantage for MOSFET-based ESD strategies. Modeling of one- and two-dimensional effects of avalanche multiplications is important for the prediction of MOSFET second breakdown in advanced technologies [20]. Boselli, Reddy and Duvvury showed that as the substrate wafer is increased from 2 to 50 Ω cm, the MOSFET second breakdown for 0.5 μm MOSFET increases from approximately 4 to 5 mA/μm [21]. As the substrate doping concentration decreases, the second derivative of the second breakdown current versus design width decreases. In the case of the higher p-substrate doping, in 2 Ω cm, there is a rapid decrease in the second breakdown current with the design width. As the substrate doping increases, $\mathrm{d}I_{t2}/\mathrm{d}W$ is reduced. This manifests itself in that as the substrate doping increases the MOSFET current constriction is reduced leading to a wider width of the current constriction, and a higher W_{eff}.

Unfortunately, there are also many ESD disadvantages of p^- wafers. These include low guard ring efficiency, low latchup tolerance, and lower diode ESD robustness. These effects can compensated by new processes and design. Latchup can be improved

by an increase in the substrate contacts, double guard rings, trench isolation, and alternative design strategies [22–26]. Lower diode ESD robustness can be achieved with alternative diode implementations by not using the substrate as a discharge path. The low-doped substrate can also be modified by p^+ heavily doped buried layers to compensate the degradation effects [22–26].

In Chapter 5, we will discuss well regions. This chapter will discuss the distinctions of diffused wells, retrograde wells, triple well and sub-collectors from a general approach. Distinctions will be drawn based on ESD experimental work showing these have significant effect on the operation of vertical diodes, vertical parasitic *pnp* transistors, lateral *npn* transistors and other ESD structures of interest. This knowledge can be used for an understanding of ESD, latchup and minority carrier injection. Samples of some experimental results will be shown of recent discoveries in ESD associated with well structures.

PROBLEMS

4.1. Given a square trench defined *p–n* diode structure on a *p*-substrate with an anode width of W_a, and a deep trench defined cathode of width W_c, whose trench depth is L_{DT}, derive the downward thermal series resistance assuming 45° conical region of current flow from the anode. Assume no heat transfer through the trench region.

4.2. In the above structure assume a shallow trench isolation (STI) region surrounding the anode of depth L_{STI}, whose anode junction depth is Lj. Assume no heat transfer through the STI isolation region. Derive the thermal resistance addressing the STI region with the same assumptions as Problem 1.

4.3. In the above structure, assume heat flows through the surface to the metal interconnects. Assume a contact width W_{ca}, and height L_{ca}. From the silicon surface, show the total thermal resistance where the interconnects are the 'up' resistance in parallel with the 'down' resistance to the substrate.

4.4. Derive the thermal resistance for a p^-/p^+ epitaxial wafer.

4.5. Derive the thermal resistance for a p^- wafer with a p^+ buried layer. Assume a gaussian profile for the p^+ buried layer.

4.6. Derive the thermal resistance for an SOI layer.

REFERENCES

1. R. Joy and E. S. Schlig. Thermal properties of very fast transistors. *IEEE Transactions on Electron Devices* 1970; **ED-17** (8): 586–593.
2. J. Melcher. *Continuum Electromechanics*, M.I.T. Press, 1982.
3. J. A. Kong. *Theory of Electromagnetic Waves*. Wiley 1975.
4. R. R. Troutman and M. J. Hargrove. Transmission line Model for latchup in CMOS circuits. *Digest of Technical Papers for the 1983 Symposium on VLSI Technology*, 1983; 56–59.
5. R. R. Troutman and M. J. Hargrove. Transmission line modeling of substrate resistance and CMOS Latchup. *IEEE Transactions on Electron Devices* 1986; 945–954.
6. R. Troutman. *Latchup in CMOS Technology: The Problem and the Cure*. Kluwer, 1984.
7. D. B. Estreich, A. Ochoa and R. W. Dutton. Analysis of latchup prevention in CMOS IC's using epitaxial buried layer process. *International Electron Device Meeting (IEDM) Technical Digest*, 1978; 230–234.

8. D. B. Estreich. *The Physics and Modelling of Latchup and CMOS Integrated Circuits*. Integrated Circuits Laboratory, Stanford University, California, November 1980.
9. R. Troutman. Epitaxial layer enhancement of *n*-well guard rings for CMOS circuits. *IEEE Transactions on Electron Devices Letters* 1983; **ED-4:** 438–440.
10. P. Cottrell, S. Warley, S. Voldman, W. Leipold and C. Long. N-well design for trench DRAM arrays. *International Electron Device Meeting (IEDM) Technical Digest*, December 1988; 584–587.
11. S. Voldman and V. Gross. Scaling, optimization, and design considerations of electrostatic discharge protection circuits in CMOS technology. *Journal of Electrostatics* 1994; **33** (3): 327–357.
12. S. Voldman, M. Marceau, A. Baker, E. Adler, S. Geissler, J. Slinkman, J. Johnson and M. Paggi. Retrograde well and epitaxial thickness optimization for shallow- and deep-trench collar merged isolation and node Trench SPT cell and CMOS Logic Technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1992; 811–815.
13. S. Voldman. Shallow trench isolation double-diode electrostatic discharge circuit and interaction with DRAM circuitry. *Journal of Electrostatics* 1993; **31**: 237–262.
14. N. K. Verghese and D. Allstot. Rapid simulation of substrate coupling effects in mixed mode integrated circuits. *Proceedings of the IEEE Custom Integrated Circuit Conference*, 1993.
15. T. Li, C. H. Tsai, E. Rosenbaum and S. M. Kang. Substrate resistance modeling and circuit level simulation of parasitic device coupling for CMOS I/O under ESD stress. *Proceedings of the EOS/ESD Symposium*, 1998; 281–289.
16. D. Walkey *et al*. Prediction of thermal resistance in trench isolated bipolar device structures. *Proceedings of the Bipolar Circuit Technology Meeting*, 1998; 207–210.
17. D. Walkey, T. Smy, C. Reimer, M. Schroter, H. Tran and D. Marchesan. 'Modeling thermal resistance in trench-isolated bipolar technologies including trench heat flow'. *Solid State Electronics* 2002; **46**: 7–17.
18. J. S. Rieh *et al*. Structural dependence of the thermal resistance of trench-isolated bipolar transistor model. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, September 2002; 100–103.
19. S. Voldman. The effect of deep trench isolation, trench isolation and sub-collector doping on the electrostatic discharge (ESD) robustness of radio frequency (RF) ESD diode structures in BiCMOS silicon germanium technology. *Proceedings of the EOS/ESD Symposium*, 2003; 214–223.
20. A. Schutz, S. Selberherr and H. Potzl. A two-dimensional model for the avalanche effect in a MOS transistor. *Solid State Electronics* 1982; **25** (3): 177–183.
21. G. Boselli, V. Reddy and C. Duvvury. Impact of scaling on the high current behavior of RF CMOS technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 229–235.
22. M. Hargrove, S. Voldman, J. Brown, K. Duncan and W. Craig. Latchup in CMOS. *Proceedings of the International Reliability Physics Symposium*, April 1998; 269–278.
23. W. Morris. CMOS latchup. *Proceedings of the International Reliability Physics Symposium (IRPS)*, May 2003; 76–84.
24. K. K Bourdelle, Y. Chen, R. A. Ashton, L. M. Rubin, A. Agarwal and W. Morris. Evaluation of high dose, high energy boron implantation into CZ substrates for epi-replacement in CMOS technology. *IEEE Transactions on Electron Devices* 2001; **ED-48**: 2043.
25. K. C. Leong, P. C. Liu, W. Morris and L. Rubin. Superior latchup resistance of high dose energy implanted p^+ buried layers. *Proceedings of the XIIth International Conference on Ion Implantation Technology*, Kyoto, Japan, 1998; 99.
26. A. Watson and S. Voldman. The influence of deep trench and substrate resistance on the latchup robustness in a BiCMOS silicon germanium technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2004; 135–142.

5 Wells, Sub-collectors and ESD

This chapter will discuss diffused wells, retrograde wells, triple wells and sub-collectors and their influence on ESD robustness of semiconductor components from a general approach. Distinctions will be drawn based on ESD experimental work, showing these have significant effect on the operation of vertical diodes, vertical parasitic *pnp* transistors, lateral *npn* transistors and other ESD structures of interest. This knowledge can be used for an understanding of ESD, latchup and minority-carrier injection.

Semiconductor devices are placed in collectors and well regions in the substrate of a semiconductor wafer. These regions have a strong influence on the ESD robustness of a semiconductor device. The doping concentration, energy, dose and perimeter structures can influence the ESD robustness of the semiconductor device.

Depending on the technology generation, semiconductors devices are formed in single-well, double-well and triple-well technology. In this chapter, we will focus on the single-well technology where the well region is of opposite doping polarity to the semiconductor substrate. Substrate regions also have an influence on the well profiles, and can modulate the doping concentration and ESD robustness. The use of these elements as resistor elements, ballasting elements will also be discussed.

Diffused wells were common in early CMOS development prior to 1994 [1,2]. In the 1980s only a few corporations utilized high-energy MeV implanters for formation of retrograde wells [3–15]. Without high-energy implanters, the majority of BiCMOS applications also formed sub-collectors prior to epitaxial deposition.

5.1 DIFFUSED WELLS

In semiconductor technologies, isolation wells were formed to allow the placement of different semiconductor transistor types in a common substrate. These isolation wells have considerable influence on the operation and reliability of the semiconductor device inside the isolation well. These isolation well serve as a region to isolate the semiconductor device from the substrate region. From an ESD and CMOS latchup perspective, these isolation wells

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

also lead to 'parasitic' devices which influence the ESD sensitivity of the device contained in the well. These parasitic devices can also serve as a means to provide ESD protection. In the case of a p^+ diffusion inside an *n*-well region, the well region forms a diode where the p^+ diffusion serves as an anode, and the well region serves as a cathode. Additionally, the well region serves as a base region forming both a vertical and lateral *pnp* transistor with the *p*-substrate serving as a collector region [1,2]. Hence the understanding of the well regions is critical to understanding the ESD and latchup response of a semiconductor component.

In semiconductor technologies, isolation wells were diffused from the semiconductor wafer surface. A diffusion source was created which was doped or implanted to a certain dose level. Using thermal processing, the dopants diffuse into the substrate where the extension into the substrate is a function of the surface dopant concentration, the diffusion coefficient of the dopant in the material, and the time of the process for the diffusion process. In the case of a diffused well, the doping concentration as a function of position can be approximated as a Gaussian-like distribution.

From the diffusion equation, the doping profile can be expressed as a function of the complementary error function distribution, where D is the diffusion constant of the element, and t is time

$$C(x,t) = C_s erfc\left(\frac{x}{2\sqrt{Dt}}\right) = \frac{2C_s}{\sqrt{\pi}}\int_{\frac{x}{2\sqrt{Dt}}}^{\infty} e^{-\nu^2}\mathrm{d}\nu$$

To get the total dopant, integration over the well can be expressed as

$$N' = \int_0^{\infty} C(x,t)\mathrm{d}x = 2\sqrt{\frac{Dt}{\pi}}C_s$$

Doping profiles (Figure 5.1) can be represented as the doping concentration N' followed by a hot process drive-in cycle, represented by a Gaussian distribution

$$C(x,t) = \frac{N'}{\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right)$$

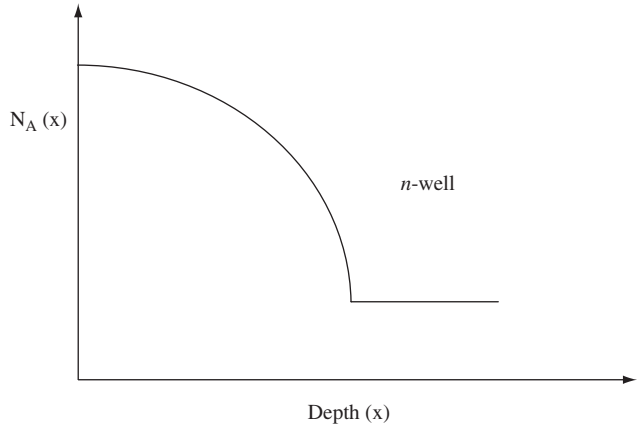

Figure 5.1 Diffused well vertical Profile

In a diffused well process, the doping gradient of the Gaussian distribution is such that it establishes a built-in electric field into the substrate region. The built-in electric field in the dopant profile establishes a drift component in the current density constitutive equation for current transport. This built-in field drift component serves in transporting holes from the emitter to the collector. The built-in drift component leads an increase in the base transport factor of the parasitic bipolar transistor element.

From the constitutive relationships of hole current and electron current

$$J = \mu_{\mathrm{h}} p_0 E_x - qD_{\mathrm{h}} \frac{\mathrm{d}p_0}{\mathrm{d}x}$$

$$J = \mu_{\mathrm{e}} n_0 E_x + qDe \frac{\mathrm{d}n_0}{\mathrm{d}x}$$

From the principle of detailed balance, these terms must balance independently. Letting the electron and hole current equal zero, the built-in field can equal

$$E_x = \frac{kT}{q} \frac{1}{p_0} \frac{\mathrm{d}p_0}{\mathrm{d}x}$$

$$E_x = -\frac{kT}{q} \frac{1}{n_0} \frac{\mathrm{d}n_0}{\mathrm{d}x}$$

This expression relates the built-in electrical field to the equilibrium electron and hole carrier concentration to the doping concentration profile. From this expression, where the hole or electron concentration is known from the doping distribution, the built-in electric field can be quantified.

These two expressions are not independent since it is well understood that the electron and hole populations are related according to the equilibrium relationship

$$n_0 p_0 = {n_i}^2$$

Taking the derivative of the equilibrium as a function of position

$$\frac{\mathrm{d}}{\mathrm{d}x}(n_0 p_0) = \frac{\mathrm{d}}{\mathrm{d}x} {n_i}^2$$

Expanding the expression

$$\frac{\mathrm{d}}{\mathrm{d}x}(n_0 p_0) = \frac{\mathrm{d}n_0}{\mathrm{d}x} p_0 + n_0 \frac{\mathrm{d}p_0}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x} {n_i}^2$$

Assuming that the intrinsic concentration is not a function of position (e.g. constant material property and temperature), then

$$\frac{1}{p_0} \frac{\mathrm{d}p_0}{\mathrm{d}x} = -\frac{1}{n_0} \frac{\mathrm{d}n_0}{\mathrm{d}x}$$

By equating the two equations for the built-in field, we can relate the gradient of the hole concentration to the gradient of the electron concentration, and obtain

$$\frac{1}{p_0}\frac{\mathrm{d}p_0}{\mathrm{d}x} = -\frac{1}{n_0}\frac{\mathrm{d}n_0}{\mathrm{d}x}$$

Since the diffused *n*-well is a function of position, its doping gradient will influence the minority-carrier diffusion of holes through this region. In the case of a vertical bipolar transistor, the diffused *n*-well region serves as a base region. Minority-carrier holes diffuse through the base region from the p^+ diffusion at the device surface. In a diffused well, the doping profile is such that the doping concentration decreases from the device surface. From the above expression, given that the gradient of the *n*-type dopant is negative, the built-in electric field is positive. The positive built-in electric field term leads to a positive hole current providing a drift component to the hole current density. The constitutive relationship, with built-in field E_{x0}, can be expressed as

$$J_\mathrm{h} = q\mu_\mathrm{h}(p_0 + p')E_{x0} - qD_\mathrm{h}\left(\frac{\mathrm{d}p_0}{\mathrm{d}x} + \frac{\mathrm{d}p'}{\mathrm{d}x}\right)$$

$$J_\mathrm{e} = q\mu_\mathrm{e}(n_0 + n')E_{x0} + qD_\mathrm{e}\left(\frac{\mathrm{d}n_0}{\mathrm{d}x} + \frac{\mathrm{d}n'}{\mathrm{d}x}\right)$$

The diffused well vertical bipolar *pnp* transistor (Figure 5.2) has the property of an increased bipolar forward current gain as a result of the built-in field assisting the hole transport. Note that the diffused well built-in field also leads to a lower reverse bipolar current gain.

Using the diffused well vertical bipolar *pnp* transistor for an ESD element provides an advantage of increased current transport when the p^+ diffusion emitter is at the device surface and the collector is the substrate [1,2].

For the diffused well bipolar transistor element, low diffused well doping concentrations leads to a low *pnp* base transistor. As the base doping concentration is decreased, the bipolar gain increases. Hence, by evaluating the diffused *n*-well sheet resistance, high sheet resistances leads to improved bipolar current gain.

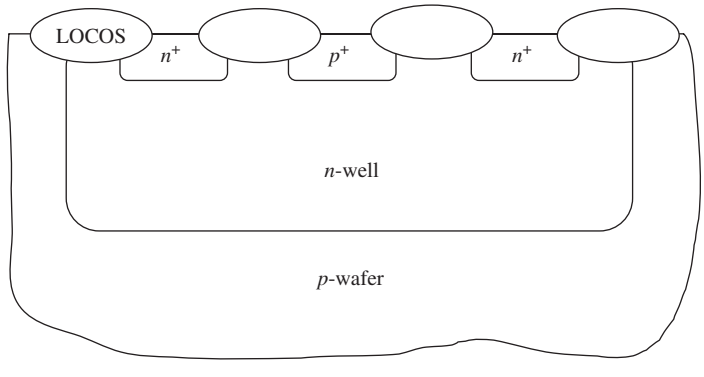

Figure 5.2 Vertical *pnp* structure

For ESD protection, this can be utilized in a *p–n* diode structure, a vertical *pnp* structure, a mixed voltage *p–n* diode string circuit, a silicon controlled rectifier (SCR), medium level trigger SCR (MLSCR), and low-voltage trigger SCR (LVTSCR). ESD protection studies by Maloney showed the advantage of the diffused well profile by providing excellent ESD results.

ESD experimental results show that as the *n*-well sheet resistance increases from 1000 Ω/sq. to 2000 Ω/sq., the vertical bipolar current gain increases leading to significant increases in the HBM and MM results [12,13].

Diffused wells, although having the advantage of transporting holes from the emitter to the collector for ESD applications, also have significant disadvantages in terms of functionality, capacitive loading, and semiconductor chip integration. First, the diffusion process has the highest doping concentration near the semiconductor device surface. This leads to an increase in capacitance at the *p*-diffusion to well metallurgical junction. Second, the high vertical bipolar gain leads to significant current injection to the substrate. This can lead to noise, and latchup concerns. From an ESD perspective, using a diffused well structure as an anode to a diode element can lead to high diode series resistance where its desired operation is diodic as opposed to a bipolar operation.

Figure 5.3 demonstrates self-heating from an ESD event in a p^+/n-well diode in a diffused well with LOCOS and STI isolation. Electrothermal simulation results shows that for the same ESD event, the diffused well process provides the lowest peak temperature in a diode structure. This is a result of the vertical profile of the diffused well.

From ESD HBM (Figure 5.4) and MM testing (Figure 5.5), at high sheet resistances, the ESD robustness of a diode structure improves. The upward region of the U-shape dependence on the *n*-well sheet resistance is the region of a diffused well profile. As the sheet resistance of the diffused well increases, the vertical bipolar gain increases, leading to discharge of current to the substrate during an ESD event.

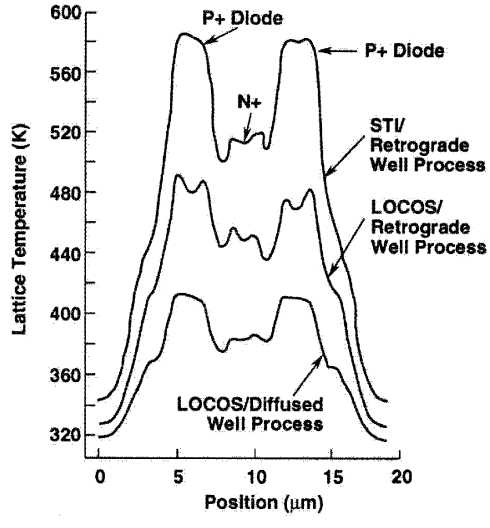

Figure 5.3 Self-heating in p^+/n-well diode highlighting LOCOS isolation in diffused versus retrograde well

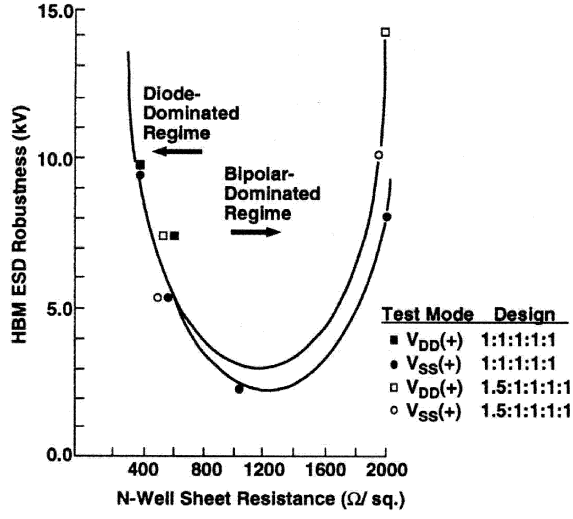

Figure 5.4 HBM ESD results as a function of *n*-well sheet resistance

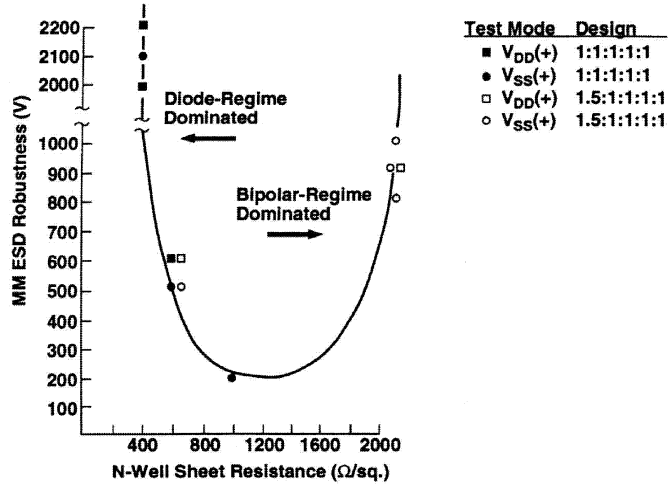

Figure 5.5 MM ESD results as a function of *n*-well sheet resistance

5.2 RETROGRADE AND VERTICALLY MODULATED WELLS

5.2.1 Retrograde Wells

In semiconductor technologies, retrograde wells (also known as vertically modulated wells) were implemented as isolation wells to allow the placement of different semiconductor transistor types in a common substrate and minimize the hot process, provide low sheet resistance wells, and reduce the risk of CMOS latchup. Retrograde wells have considerable influence on the operation, performance, latchup and ESD robustness of a semiconductor device [3–15]. Figure 5.6 is an example of a retrograde well profile.

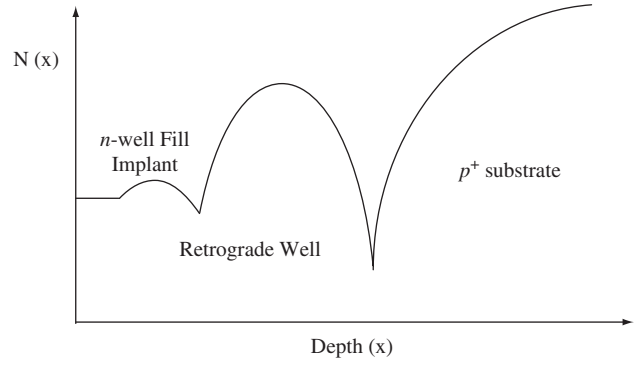

Figure 5.6 Retrograde well vertical profile

From an ESD and CMOS latchup perspective, the retrograde wells serve as the base region of the 'parasitic' devices which influence the ESD sensitivity of the device contained in the well. The parasitic device also can serve as a means to provide ESD protection. In the case of a p^+ diffusion inside a retrograde *n*-well region, the well region forms a diode where the p^+ diffusion serves as an anode, and the retrograde well region serves as a cathode. Additionally, the well region serves as a base region forming both a vertical and lateral *pnp* transistor with the *p*-substrate serving as a collector region.

In semiconductor technologies, in contrast to the diffused well, the retrograde well is not diffused from the semiconductor wafer. A retrograde well is implanted using high-energy implanters below the semiconductor wafer surface. The retrograde well can consist of a single implant, or a plurality of implants of different doses and energies. Typically a low dose and energy is placed near the surface isolation and p^+ diffusion implant to minimize p^+/well junction capacitance, yet prevent device punch-through, insure adequate implant dose under the isolation structure and adequate doping for the MOSFET threshold body effect. The advantage of the vertically modulated well profile is that the well engineering can be established to provide the optimum device design point to achieve these objectives. These lower dose implants are at times referred to as 'retrograde fill implants.' Additionally, a heavily doped retrograde implant is placed deep away from the semiconductor device surface. In this case, the implant is a two-sided Gaussian profile whose distribution is a function of the species, implant dose, and implant energy. Using thermal processing, the dopants are activated and diffuse into the substrate where the extension toward the surface and into the substrate is a function of the substrate dopant type and species, the diffusion coefficient of the dopant in the material, and the diffusion process time. In the case of a retrograde well, the doping concentration as a function of position can be approximated as a two-sided Gaussian distribution and a low-level region and lower dopant region (the low-doped, low-energy implant or plurality of implants).

In a retrograde well, the variation in doping concentration establishes a built-in electric field. The doping gradient of the Gaussian distribution typically has two polarities where the built-in electric field is opposite from the peak retrograde well implant doping concentration. In the retrograde well, the built-in electric field in the dopant profile establishes a drift component in the current density constitutive equation

for current transport. This built-in field drift component impedes the transport of minority-carrier holes from the emitter to the collector. The built-in drift component leads to a decrease in the base transport factor of the parasitic bipolar transistor element where the p^+ implant at the surface is the emitter of a vertical bipolar transistor.

From the constitutive relationships of hole current and electron current

$$J_{\mathrm{h}} = \mu_{\mathrm{h}} p_0 E_x - qD_{\mathrm{h}} \frac{\mathrm{d}p_0}{\mathrm{d}x}$$

$$J_{\mathrm{e}} = \mu_{\mathrm{e}} n_0 E_x + qD_e \frac{\mathrm{d}n_0}{\mathrm{d}x}$$

From the principle of detailed balance, these terms must balance independently. Letting the electron and hole current equal zero, the built-in field can equal

$$E_x = -\frac{kT}{q}\frac{1}{n_0}\frac{\mathrm{d}n_0}{\mathrm{d}x}$$

From this expression, where the hole or electron concentration is known from the doping distribution, the built-in electric field can be quantified. Assuming that the intrinsic concentration is not a function of position (e.g. constant material property and temperature), then

$$\frac{1}{p_0}\frac{\mathrm{d}p_0}{\mathrm{d}x} = -\frac{1}{n_0}\frac{\mathrm{d}n_0}{\mathrm{d}x}$$

Since the retrograde n-well is a function of position, its doping gradient will influence the minority carrier diffusion of holes through this region. In the case of a vertical bipolar transistor, the diffused n-well region serves as a base region. Minority carrier holes diffuse through the base region from the p^+ diffusion at the device surface. In a retrograde well, the doping profile is such that the doping concentration has a weak gradient from the n-well fill implant and increases from the device surface. From the above expression, given that the gradient of the n-type dopant is positive, the built-in electric field is negative. The negative built-in electric field term leads to a negative hole current providing a drift component to the hole current density. The constitutive relationship, with built-in field E_{x0}, can be expressed as

$$J_{\mathrm{h}} = q\mu_{\mathrm{h}}(p_0 + p')E_{x0} - qD_{\mathrm{h}}\left(\frac{\mathrm{d}p_0}{\mathrm{d}x} + \frac{\mathrm{d}p'}{\mathrm{d}x}\right)$$

$$J_{\mathrm{e}} = q\mu_{\mathrm{e}}(n_0 + n')E_{x0} + qD_{\mathrm{e}}\left(\frac{\mathrm{d}n_0}{\mathrm{d}x} + \frac{\mathrm{d}n'}{\mathrm{d}x}\right)$$

After the minority-carrier holes are transported after the peak retrograde well doping peak, the built-in electric field polarity is of comparable magnitude, but of reverse polarity. The minority-carrier holes are then assisted to flow to the n-well-to-substrate metallurgical junction.

The retrograde well vertical bipolar *pnp* transistor has the property of an decreased bipolar forward current gain as a result of the built-in field impeding the minority-carrier hole transport. Depending on the substrate doping concentration, there is an increase in the reverse bipolar current gain where the substrate serves as the emitter and the surface *p*-diffusion serves as a collector.

A significant factor that influences the forward and reverse bipolar current gain is the energy and dose of the retrograde *n*-well (Figure 5.7). Since the implant dose and energy is not limited to the surface semiconductor device optimization, the retrograde well implant allows for significant design freedom and freedom of action. The base width can be broadened by using higher energy implants, moving the implant from the device surface. The recombination in the base region can be increased by increasing the n-type dopant concentration. As the retrograde well dose is increased, the forward bipolar current gain is decreased. As the retrograde well dose in increased, the vertical bipolar current gain decreases. This decrease the injection of minority carrier holes into the p^- or p^+ substrate region. This decreases the effectiveness of an ESD element when utilized as a vertical bipolar *pnp* transistor.

As the retrograde well dose increases, the sheet resistance of the *n*-well region is decreased. With increasing well dose, the diode formed by the p^+ diffusion as the anode, and the retrograde well as the cathode, the diode series resistance decreases. As the diode series resistance decreases, the Joule heating in the region decreases in the well region. Additionally, as the doping concentration is increased, the intrinsic temperature in the retrograde well region increases, providing an improved thermal stability.

Additionally, as the retrograde well doping concentration increases, the doping concentration under the device isolation is also increased if the well is sufficiently close to the semiconductor isolation structures. As the retrograde well dose increases, the peak local temperature under the isolation decreases during high-current and ESD events.

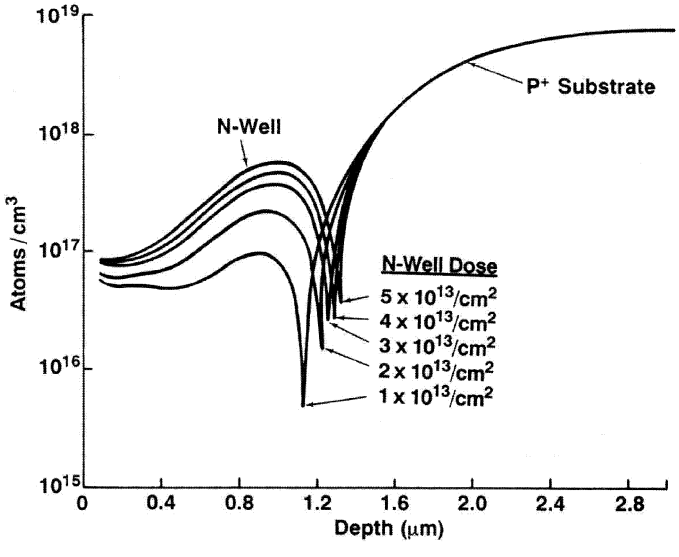

Figure 5.7 Retrograde well vertical profile as a function of MeV implant dose

With retrograde well structures, the operation of silicon-controlled rectifiers (SCR), medium level SCR (MLSCR), and low-voltage trigger SCR (LVTSCR) are significantly affected as a result of the *pnp* bipolar current gain reduction.

5.2.2 Retrograde Well Substrate Modulation

Retrograde wells formed in a p^{++} heavily doped substrate with an epitaxial region, or abutting a heavily doped buried layer implant can undergo modulation of the retrograde well profile (Figure 5.8). This effect is known as the retrograde well substrate modulation effect [3–5]. This effect has considerable influence on the ESD protection of a p^{+}/n-well diode network.

The retrograde well profile can be represented as a Gaussian implant at x_{w0} and n-well fill doping profile,

$$N_{\mathrm{D}}(x) = N_{\mathrm{fill}} + N_{\mathrm{D}} \exp\left\{ -\left[\frac{x - x_{w0}}{\sqrt{2\sigma}}\right]^2 \right\}$$

We can define an effective epitaxial thickness as

$$T_{\mathrm{eff}} = T_{\mathrm{epi}} + \lambda$$

with the effective thickness determined by the epitaxial flat zone at the end of the surface implant (or p-well), and the beginning of the p^{+} substrate (or p^{++} buried layer). In the p^{++} substrate model case, as discussed in the previous chapter, an effective depth factor λ is to address the extra effective thickness between the epitaxial flat zone and the heavily doped flat zone of the p^{++} wafer to add an increase in the epitaxial effective width.

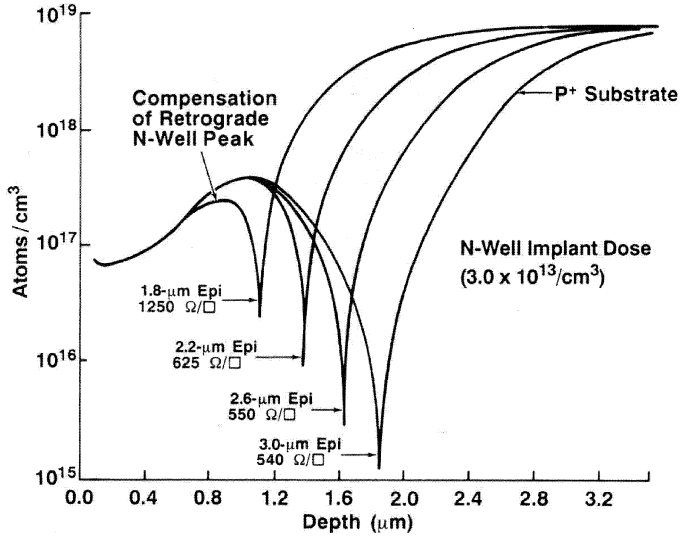

Figure 5.8 Retrograde well vertical profile as a function of epitaxial thickness

For the p^{++} wafer, a concentration profile can be assumed to be equal to

$$N_{\text{sub}}(x) = \frac{N_{\text{sub}}}{2} erfc\left[\frac{x - x_{\text{e}}}{x_0}\right]$$

where we determine the epitaxial depth position from the above expression, letting $x_2 = x_{\text{e}}$. From this we can express the effective depth parameter as

$$\lambda = \frac{x_0}{2} erfc^{-1}\left\{\frac{2N_{\text{epi}}}{N_{\text{sub}}}\right\}$$

As the epitaxial thickness T_{epi} decreases, the flat zone of the *p*-region decreases. Variations in the epitaxial thickness is common due to the control of incoming wafers. With the variation, the width term λ does not change, but is shifted toward the device surface. Additionally, the MeV retrograde well implant also is implanted deeper toward the p^{++} substrate. The MeV implant dose is significantly lower than the p^{++} substrate wafer, leading to strong modulation of the profile, the capacitance and the *n*-well sheet resistance.

Given that an element is placed in the well, the ESD robustness of the device has been shown to vary strongly with the *n*-well sheet resistance variation. For example, an STI-bound p^{+}/n-well diode structure is a function of the diode series resistance and the vertical bipolar current gain. As the substrate impinges on the *n*-well region, the physical well depth decreases and the *n*-well sheet resistance increases. For a diode structure, this leads to a higher diode series resistance. Additionally, the vertical bipolar *pnp* current gain increases as a result of the smaller base width and the reduction of the Gummel number (e.g. integral of the dopants in the base region). Experimental results have shown that the *n*-well sheet resistance can vary from 700 to 1500 Ω/sq. and the vertical bipolar gain can increase from $\beta = 2$ to 8 [4,5,7–10,14].

As the epitaxial thickness is varied, for a fixed well dose and energy implant, experimental results show the HBM ESD results decrease with increasing well sheet resistance from 300 to 1000 Ω/sq. from roughly 8 to 3 kV (Figure 5.9). With the decrease in the epitaxial thickness, the effective film thickness of the series resistance decreases as well as the total doped elements. In the self-heating of the resistor element, the film thickness as well as the doping concentration play a role [8].

The ESD robustness linearizes when the data is plotted as a function of the square root of the well sheet resistance over the film thickness (Figure 5.10), roughly

$$V \propto \frac{\sqrt{t_{\text{film}}}}{\sqrt{\rho_{\text{well}}}}$$

Evaluating of the *n*-well doping profiles, as the epitaxial region is too thin, the peak *n*-well concentration becomes consumed by the p^{++} substrate region, significantly changing the doping profile in the *n*-well. Although the base width is decreasing and the vertical bipolar current gain is increasing to approximately 8, ESD results decrease due to the high series resistance (serving as the diode series resistance and the vertical bipolar base resistance). From the work of Maloney, when the vertical bipolar current gain is less than 10, the effectiveness of the vertical bipolar element is not strong during ESD events [1,2].

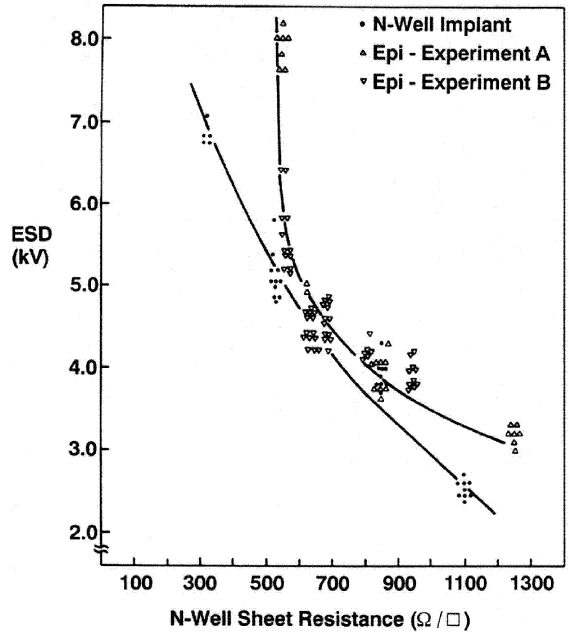

Figure 5.9 HBM ESD and sheet resistance as a function of well and epitaxial variation splits

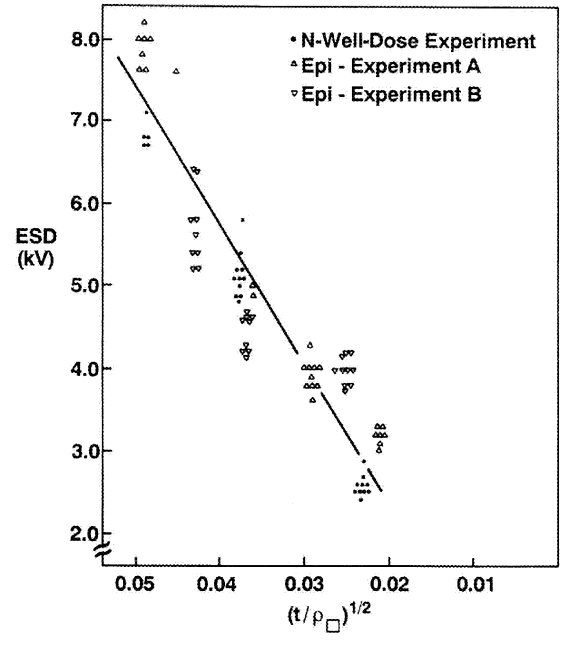

Figure 5.10 HBM ESD results compared with normalized metric (e.g. film thickness and sheet resistance)

5.2.3 Retrograde Wells and ESD Scaling

ESD robustness of a diode structure in a retrograde well is a strong function of the retrograde well sheet resistance, and profile [5–16]. In the case of a p^+ diffusion inside a retrograde n-well region, the well region forms a diode where the p^+ diffusion serves as an anode, and the retrograde well region serves as a cathode. In this diode representation, the STI-bound diode forms a metallurgical junction with the low-doped region of the n-well. The n-well serves as a cathode resistive element in series with the metallurgical junction.

Assume in the retrograde well, that the failure of the diode structure is related to the volume inside the retrograde well between the metallurgical junction, and the n^+ well contact region [8]. Assume that there exists stored energy balanced by the Joule heating

$$\mathrm{d}E = C_{\mathrm{p}}(T)\mathrm{d}T = I^2(t)R(T)\mathrm{d}t$$

where $C_{\mathrm{p}}(T)$ is the heat capacity, $I(t)$ is the current pulse, $R(T)$ is the resistance as a function of temperature in the well region. Maloney showed that this can be put in a temperature formulation or an energy formulation. In an energy representation, let the resistance be represented as $R(T) = R_0\, g(E)$ where $g(E)$ is a functional of energy. Let the heat capacity $C_{\mathrm{p}}(T) = C_{\mathrm{p}}\, h(E)$ where $h(E)$ is a functional.

$$\mathrm{d}E = I^2(t)R_0 g(E)\mathrm{d}t$$

Separating the variables and integrating over both sides of the equation

$$\int \frac{\mathrm{d}E}{g_0(E)} = R_0 \int_0^{\tau} I^2(t)\mathrm{d}t$$

Note from this equation, the left-hand side is a function of energy, and the right-hand side is a function of the resistance integrated over the square of the current for a given time τ. We can create a functional F, where the F is a constant.

$$F = \int \frac{\mathrm{d}E}{g_0(E)} = R_0 \int_0^{\tau} I^2(t)\mathrm{d}t$$

The functional F is equal to the integration of the energy over the length of the pulse. Assuming that the net increase in current beyond the ESD time constant τ, the integral can be extended over all time. To relate the internal energy deposition to the ESD pulse and external variables, one can relate an average resistance which is the effective resistance of the pulse source, and the internal resistance where the pulse is flowing.

$$R_{\mathrm{AVG}} \int_0^{\infty} I^2(t)\mathrm{d}t = \frac{CV^2}{2}$$

Then

$$F = \int \frac{\mathrm{d}E}{g_0(E)} = R_0 \frac{CV^2}{2}\left(\frac{1}{R_{\mathrm{AVG}}}\right)$$

Treating the well region as a single sheet resistance term, we can model the lateral retrograde well where it is the bulk resistivity ρ, width W, cross-sectional thickness t, and mass density ρ_{d}

$$R_0 = \frac{\rho}{W^2 t^2 \rho_{\mathrm{d}}}$$

Substituting in for the value of R_0 internal resistance,

$$F = \int \frac{\mathrm{d}E}{g_0(E)} = \left(\frac{\rho}{W^2 t^2 \rho_{\mathrm{d}}}\right)\frac{CV^2}{2}\left(\frac{1}{R_{\mathrm{AVG}}}\right)$$

In this form, we can relate the voltage on the source voltage V, and the internal dimensions of the internal resistance element. If we assume that the function F is fixed and is associated with the ESD failure, then if the failure mechanism is the same, the right-hand side of the equation must equal a constant. In this fashion, we can express the system as a function of two different voltage conditions or rearranging, we can express this as

$$\frac{V'^2}{V^2} = \frac{\left(W'^2 t'^2 \rho'_{\mathrm{d}}/\rho'\right)}{\left(W^2 t^2 \rho_{\mathrm{d}}/\rho\right)}\frac{\left(2R'_{\mathrm{AVG}}/C'\right)}{\left(2R_{\mathrm{AVG}}/C\right)}$$

Hence, assuming the same average resistance, and the same capacitor source, we can simplify the expression as a function of the design and material parameters. In this fashion, we can relate the HBM voltage on the source to the characteristics of the process.

Assuming the same source, and the same mass density, then we can show that this can be represented as

$$\frac{V'^2}{V^2} = \frac{\left(W'^2 t'^2/\rho'\right)}{\left(W^2 t^2/\rho\right)}$$

In this form, we can now relate the sheet resistance, thickness and the design width to the HBM ESD robustness. If we assumed that in MOSFET constant electric field scaling that dimensional scaling is preserved, then

$$W' = W/\alpha$$

$$t' = t/\alpha$$

$$\rho' = \rho/\alpha$$

and ESD scaling with the scaling of the well will follow the relationship of

$$V' = \frac{V}{\alpha^{3/2}}$$

Hence if the structure was scaled in width, film thickness and resistance, the ESD robustness of a diode structure would decrease with the scaling parameter α.

Assuming that a design is fixed in width, and hence the only scaling relationship involves well depth and sheet resistance. We can then establish the retrograde well scaling under these constraints as

$$\frac{V'^2}{V^2} = \frac{(t'^2/\rho')}{(t^2/\rho)}$$

where t is the cross-sectional thickness of the retrograde well region laterally and ρ is the sheet resistance. Let us define the depth t as the depth of the well under the isolation region

$$\frac{V'}{V} = \left(\frac{L'_{\text{well}}}{L_{\text{well}}}\right)\left(\frac{\rho}{\rho'}\right)$$

Given that the dimensional scaling is such that

$$L'_{\text{well}} = L_{\text{well}}/\alpha$$

and from MOSFET constant electric field scaling

$$N' = N\alpha$$

where N is the doping concentration, then the scaling relationship is

$$\frac{V'}{V} = \frac{1}{\alpha^{1/2}}$$

This scaling relationship for the n-well states that as the dimensional scaling occurs, the ESD results will decrease according to the above scaling parameter. To compensate this effect, ESD scaling theory for retrograde wells can be developed to preserve the ESD protection level with scaling. Hence, a scaling parameter can be established which has the following modification to the standard MOSFET electric field scaling relationship

$$N' = N\gamma\alpha$$

$$L_{\text{well}'} = L_{\text{well}}/\alpha$$

Hence, to prevent the ESD robustness decreasing with retrograde well scaling the following relationship must be true

$$\frac{V'}{V} = \left(\frac{L_{\text{well}'}}{L_{\text{well}}}\right)\left(\frac{\rho}{\rho'}\right) = 1$$

where

$$\frac{V'}{V} = \left(\frac{\gamma}{\alpha}\right)^{1/2} = 1$$

By setting the two scaling variables as equal, ESD robustness does not decrease with the MOSFET constant electric field scaling theory. This 'constant ESD scaling' was achieved under the framework of MOSFET constant electric field scaling theory where an additional well scaling parameter was defined [8].

With technology scaling of the well depth, the ESD robustness will decrease for a given retrograde well profile (Figure 5.11).

5.3 TRIPLE-WELL AND ISOLATED MOSFETs

In applications where it is an advantage to isolate both the *n*- and *p*-channel MOSFETs from the chip substrate, triple-well technology can be utilized. First, triple-well technology allows circuits to be independent of the substrate bias conditions. This is an advantage for noise isolation, mixed signal and dynamic threshold MOSFET applications. In BiCMOS applications, many substrates are biased negatively, and triple-well, or 'isolated MOSFETs' are used to allow independent biasing of the channel region. In advanced CMOS, triple wells are used to isolate the *n*-channel MOSFET from noise. Additionally, allowing independent biasing of the substrate allows for biasing of the channel region in the isolated region, and providing a different back-bias condition, or dynamic threshold MOSFETs (DTMOS) devices.

Triple-well structures (Figure 5.12) influence the α-particle sensitivity, latchup and electrostatic discharge protection. From the electrostatic discharge protection, there are multiple issues that occur with the addition of a triple-layer region.

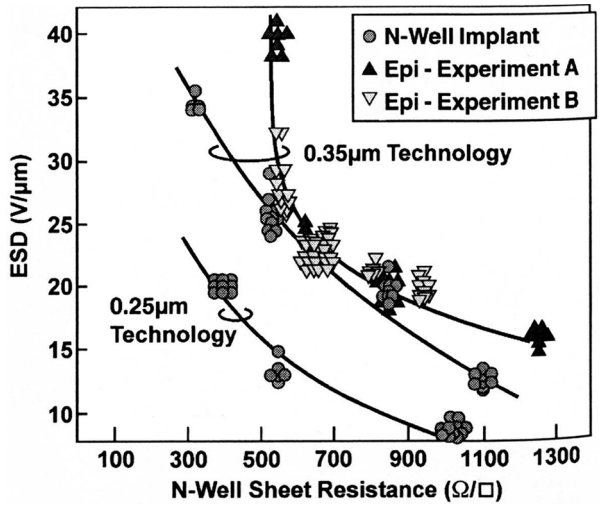

Figure 5.11 HBM ESD and *n*-well sheet resistance as a function of technology generation

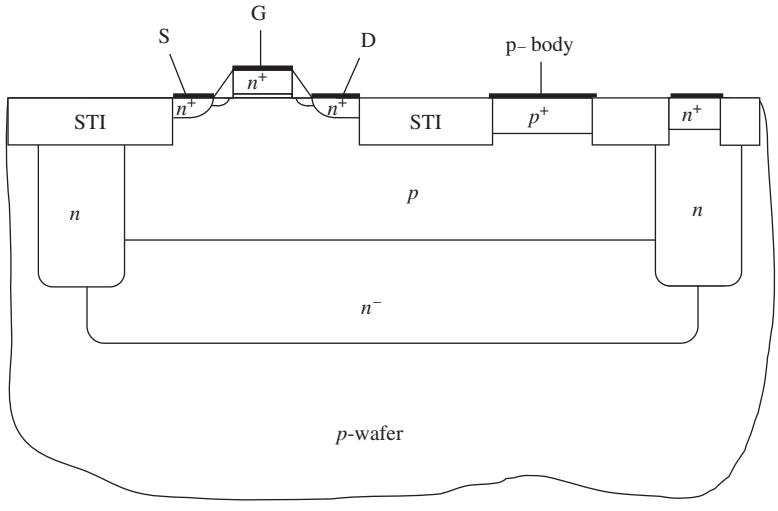

Figure 5.12 Triple-well structure

First, the inclusion of a doped layer of polarity opposite to that of the substrate region introduces two additional metallurgical junctions, with one formed between the epitaxial region and the buried layer, and a second between the buried layer and the substrate region. The addition of this layer forms a vertical *npn* parasitic bipolar transistor.

Second, the inclusion of the doped layer forms an isolated 'epitaxial layer' which has a higher series resistance compared with the substrate. The high epitaxial resistance will influence the turn-on condition for the MOSFET snapback and second breakdown physics. Given that the isolated epitaxial region has a higher resistance, the MOSFET snapback voltage will be lower. The current that flows from the MOSFET will discharge to the isolated epitaxial region contact, or the triple-well buried layer. From the current model, avalanche generation can be expressed as a function of the multiplication factor and the total current flowing through the high electric field region.

We can express the generation current as

$$I_G = (M - 1)(I_{DS} + I_C)$$

When the electric fields are low, M is unity, leading to no generation occurring. As the electric field increases in the transistor, avalanche multiplication increases, leading to the generation term. In this expression, some of the hole generation current will serve as base current for the lateral bipolar element, while some of the holes generated will enter the isolated epitaxial region, serving as epitaxial region current. The epitaxial current can be expressed as

$$I_{epi} = (M - 1)(I_{DS} + I_C) - I_B$$

The epitaxial current in a MOSFET is typically an issue for hot electron generation and total power consumption. For ESD analysis, this epitaxial current is important

because it is associated with the voltage drop that occurs locally in the MOSFET region at high currents.

$$V_{\text{iso}} = I_{\text{epi}} R_{\text{iso}}$$

Hence, we can express the voltage drop in the isolated region as

$$V_{\text{iso}} = [(M-1)(I_{\text{DS}} + I_{\text{C}}) - I_{\text{B}}]R_{\text{iso}}$$

When the voltage drop equals the forward bias voltage of the *p–n* junction formed between the source and the isolated region, MOSFET snapback occurs. We can express this as the condition

$$V_{\text{BE}} = V_{\text{iso}} = [(M-1)(I_{\text{DS}} + I_{\text{C}}) - I_{\text{B}}]R_{\text{iso}}$$

or a condition for MOSFET snapback occurs when

$$V_{\text{BE}} \geqslant [(M-1)(I_{\text{DS}} + I_{\text{C}}) - I_{\text{B}}]R_{\text{iso}}$$

An advantage of this condition is the MOSFET snapback can be altered independent of the starting substrate wafer. The isolated resistance can be obtained from the sheet resistance, the isolated region contact, and the placement of the isolating triple-well region. This isolated region is influenced by the depth of the triple-well isolating region, and its bias dependence. Consideration of the well bias and its modulation must be addressed in the integration of the MOSFET in the triple-well structure.

For ESD diode implementations, there are distinctly different issues. ESD diode implementation which is used to discharge current to the *p*-substrate must be modified due to the isolation from the chip substrate. For example, *n*-well-to-substrate diodes must be modified to address the presence of the buried layer region. As a result, novel ESD design schemes must be implemented in order to provide ESD protection with these diode structures.

5.4 BALLAST RESISTORS

Using resistor ballasts for ESD networks and peripheral circuits is a common practice in semiconductor chips. It is possible to use *n*-wells to provide these ballast resistors by direct integration with a structure, or as an independent structural element. Using *n*-well structures as ballast resistors has significant advantage because of the geometrical structure and the doping concentration. A first advantage of using a well structure as a ballast resistor is that it does not require special masking or silicide removal. In STI-defined technology, the resistor is formed under the STI isolation away from the semiconductor surface. STI-defined n^+ diffusions serve to provide contacting to the buried resistor element. The STI region over the resistor prevents silicidation of the resistor element. A second advantage of the *n*-well resistor is that the width of the resistor element is such to avoid high current density. In surface diffusion resistors, the thin dopant film can lead to a high current density and self-heating. A third

advantage of the *n*-well resistor is that the velocity saturation current density is low, allowing for early current saturation. To evaluate a well resistance, from the thermal diffusion equation, it can be shown that the concentration from the formation of the well can be represented as

$$C(x,t) = C_s erfc\left(\frac{x}{2\sqrt{Dt}}\right) = \frac{2C_s}{\sqrt{\pi}}\int_{\frac{x}{2\sqrt{Dt}}}^{\infty} e^{-\nu^2}\mathrm{d}\nu$$

The total dopants can be found by integration over the profile

$$N' = \int_0^{\infty} C(x,t)\mathrm{d}x = 2\sqrt{\frac{Dt}{\pi}}C_s$$

After the hot process of time t, the well can be represented with a doping concentration

$$C(x,t) = \frac{N'}{\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right)$$

In a nonuniform profile, the conductance of the well resistor can be found by integration over the well region from the top surface ($x = 0$) to the metallurgical junction, $x = X_W$

$$G = \frac{W}{L}\int_0^{X_W} q\mu n(x)\mathrm{d}x$$

When the background doping concentration is negligible compared with the well implant dose, we can solve for the conductance of the well from the Gaussian representation

$$G = \frac{N'q}{\sqrt{\pi Dt}}\frac{W}{L}\int_0^{X_W} \mu_n\left[\exp\left(-\frac{x^2}{4Dt}\right)\right]\mathrm{d}x$$

In semiconductor devices, the depth of the isolation can be of the same order of magnitude as the well depth in advanced technologies. In shallow trench isolation, the depth of the isolation can be represented as

$$G \simeq \frac{N'q}{\sqrt{\pi Dt}}\frac{W}{L}\int_{X_{STI}}^{X_W} \mu_n\left[\exp\left(-\frac{x^2}{4Dt}\right)\right]\mathrm{d}x$$

In the case of a retrograde well, dopants are implanted at a depth below the surface, and a second well implant or background 'fill' implant is formed. Assuming a single-well implant, where the fill implant is significantly less than the high-energy MeV implant we can approximate the conductance as this can be expressed as

$$G \simeq \frac{W}{L}\int_{X_{STI}}^{X_W} qN_0\mu_0 + \frac{N'q}{\sqrt{\pi Dt}}\mu_n\left[\exp\left(-\frac{(x-x_0)^2}{4Dt}\right)\right]\mathrm{d}x$$

The saturation current of an *n*-well resistor can be expressed as

$$J_{\text{sat}} = qN_{\text{well}}\nu_{\text{sat}}$$

where the saturation current is equal to the product of the charge of the electron, the *n*-well doping concentration, and the saturation velocity of the electron in the well region.

Since the *n*-well doping concentration is significantly lower than typical diffusion doping concentration, the saturation current is significantly lower using *n*-well resistors. As a result of the saturation phenomenon, an *n*-well resistor will have an *I–V* characteristic (Figure 5.13) exhibiting a linear region at low voltages. As the voltage across the resistor increases, the *I–V* characteristic will plateau and be voltage independent over a range of voltages. This provides a resistor with a linear region for functional use, and a plateau region which limits the current through the structure. At high voltages, avalanche breakdown can occur if the voltage between the *n*-well and the substrate exceeds the avalanche multiplication voltage. In this case, snapback occurs in the resistor element. Additionally, if self-heating occurs prior to avalanche snapback, a negative resistance regime can initiate prior to avalanche breakdown. This is dependent on the geometric properties, and thermal characteristics of the *n*-well resistor structure compared with the doping profile of the *n*-well and *n*-well-to-substrate junction.

In the case of a retrograde well, the nature of the characteristic is a function of the doping profile and the usage of '*n*-well fill implants'; *n*-well fill implants can modulate the resistance from the n^+ contact to the heavily dosed retrograde well implant, and *n*-well resistors with low-doped fill implants will have a more resistive linear regime in the *I–V* characteristic and softer transition into the 'saturated regime.'

In a resistor structure, electric fields can approach the velocity saturation level of an electron. A form of the velocity–electric field relationship can be expressed as

$$\nu_{\text{d}}(E) = \nu_{\text{sat}} \frac{(E/E_{\text{c}})}{\left[1 + (E/Ec)^{\beta}\right]^{1/\beta}}$$

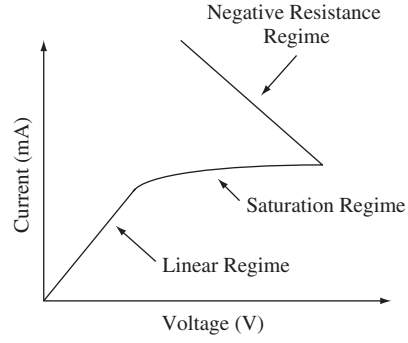

Figure 5.13 *I–V* characteristic of *n*-well resistor

where E is the electric field, E_c is the critical electrical field, and β is a parameter. Substitution into the drift equation, we obtain

$$J(E) = qn\nu_{sat}\frac{(E/E_c)}{\left[1+(E/E_c)^{\beta}\right]^{1/\beta}}$$

At high electric fields, the internal electric field approaches the critical electric field. At the critical electric field, roll-off of the current density occurs, with the saturated current density can be expressed as

$$J_{sat}(E > E_c) \simeq qn\nu_{sat}$$

where the doping concentration is the concentration of the well resistor element.

Resistor elements will undergo three regions of operation. In the linear regime, the device has a linear resistance characteristic. As the device approaches the critical electric field, the resistor enters a saturation region. Expressing the equation consistent with velocity saturation and avalanche models [17], let us express the current voltage regimes as

Linear regime

$$I = \frac{V}{R_{lin}}$$

Saturated regime

$$I = \frac{V}{R_{lin}\sqrt{1+\left(\frac{V}{V_c}\right)^2}}$$

Avalanche regime

$$I = \frac{V}{R_{lin}\sqrt{1+\left(\frac{V}{V_c}\right)^2}}M$$

$$M = \frac{1}{1-A\exp\left(-\frac{B}{V_d}\right)}$$

where A, B are physical constants,

$$V_d = \theta V$$

$$\theta = 1 - \frac{E_x L}{V}$$

and

$$1/R_{\text{lin}} \simeq G \simeq \frac{N'q}{\sqrt{\pi Dt}} \frac{W}{L} \int_{X_{\text{STI}}}^{X_{\text{W}}} \mu_n \left[\exp\left(-\frac{x^2}{4Dt} \right) \right] \mathrm{d}x$$

where E_x is the location at which the space charge density transitions into the avalanche multiplication, L is the length of the resistor element, and V is the voltage across the resistor. In this development, the linear resistance is related to the retrograde well implant. The total conductance is obtained by integrating over the retrograde well implant profile from the shallow trench isolation to the well–substrate junction region.

In a retrograde well process, the linear region is a strong function of the n-well fill implant (Figure 5.14). The low-dose, low-energy fill *n*-well implant has a strong effect on the linear portion of the ballast resistor model. It also plays a role in the slope of the saturation regime. Experimental results show that as the *n*-well fill dose increases, the saturation region has a more abrupt transition compared to the non-fill implant case. The saturation region has a lower $\mathrm{d}I/\mathrm{d}V$ dependence as well.

5.5 SUB-COLLECTORS

Sub-collectors have considerable influence on the operation, performance, latchup and ESD robustness of a semiconductor device [18–21]. In semiconductor technologies, sub-collectors are traditionally formed as the collectors of a bipolar transistor element. Sub-collectors can also serve as a heavily doped implant for a well region, or can serve as an isolating implant. Sub-collectors can be formed using a diffusion process or implanted. In the case of a diffusion/epitaxial sub-collector process, the sub-collector is formed using a diffusion source and driven into low-doped substrate region. The sub-collector characteristic is a function of the dopant dose, species and hot process cycle. Typical sub-collectors are formed using arsenic, phosphorus and antimony dopant types. The dopant is implanted into a surface region, followed by an epitaxial growth over the sub-collector, followed by a hot process step. In this method, the

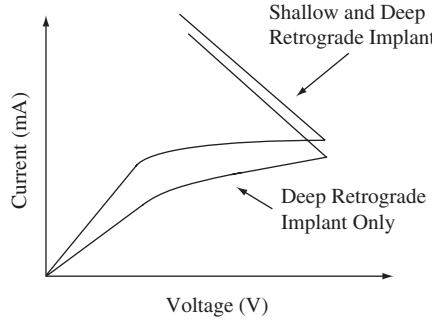

Figure 5.14 *I–V* characteristic of *n*-well resistor for different well profiles (with and without the *n*-well implant)

doping concentration can be significantly high compared with standard CMOS diffused or retrograde well technology. In an implanted sub-collector, the dose of the sub-collector is significantly reduced to the levels of a CMOS retrograde well implant (Figure 5.15).

A distinction of epitaxially grown sub-collectors is that the doping concentration is such that the electron–hole pair recombination time is small and sheet resistances are low compared with a retrograde well. The recombination time is Auger recombination time dominant as opposed to Shockley–Hall–Read dominant. CMOS diffused wells and retrograde well doses are such that the recombination time is Shockley–Hall–Read dominant. In this case, the diffusion length of the minority carrier hole is on the same scale as the sub-collector width. As a result, the *pnp* vertical bipolar current gain formed by the *p*-base region, sub-collector and substrate is recombinant dominated. A second aspect of a sub-collector structure has a sheet resistance significantly lower than the CMOS retrograde well.

Sub-collector doping concentration and profiles can have an influence on silicon bipolar transistor devices. Two key figures of merit for a bipolar transistor device are the unit current gain cutoff frequency f_T and unit power gain cutoff frequency f_{MAX}. The collector structure influences the unit current gain cutoff frequency f_T as a result of the Kirk effect. At high frequencies, the doping concentration of the collector region and pedestal implant influences the point at which the peak f_T is reached. The unit current gain cutoff frequency f_T is inversely related to the emitter-to-collector transit time τ_{EC}.

$$\frac{1}{2\pi f_T} = \tau_{EC} = \tau_E \tau_B + \tau_{CSL} + \tau_C$$

where

$$\tau_E = \frac{C_{eb} + C_{bc}}{g_m}$$

$$\tau_B = \frac{W^2{}_B}{KD_B}$$

$$\tau_{CSL} = r_C(C_{c-sx} + C_{bc})$$

$$\tau_C = \frac{x_C}{v_{sat}L}$$

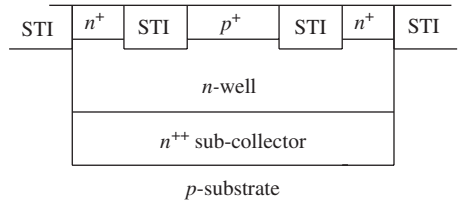

Figure 5.15 STI-bound p^+/n-well diode with sub-collector implant

The emitter transit time is a function of the emitter–base capacitance, base–collector capacitance, and transconductance. The base transit time is a function of the base width, and diffusion coefficient and base grading factor, K. The collector terms consists of the collector capacitance, collector resistance and velocity saturation term.

The unit power gain cutoff frequency f_{MAX} is a function of the unit current gain cutoff frequency, base resistance, collector–base capacitance, and collector-to-substrate capacitance. The substrate sidewall is either silicon or deep trench defined. Minimizing the collector–base capacitance C_{cb}, collector-substrate C_{csx}, and base resistance, R_b improves f_{MAX}.

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi R_b C_{bc}}}$$

From the Johnson limit formulation, where

$$V_m f_T = E_m \nu_s / 2\pi$$

The maximum voltage V_m, can be the BV_{CEO} of the bipolar device, demonstrating an inverse relationship between the breakdown voltage and the frequency response of the transistor. The sub-collector doping concentration has a significant effect on the frequency response and the breakdown voltage of the transistor.

From the experimental results on retograde *n*-wells and diffused well structures, ESD results showed that as the *n*-well sheet resistance increased from 300 to 1000 Ω/sq., the ESD robustness of a diode-based structure decreased [5–15]. This was achieved by varying the *n*-well dose, energy and epitaxial thickness. It was also shown in a diffused well that that as the *n*-well sheet resistance increased from 1000 to 2000 Ω/sq., the ESD robustness improved [12,13]. The ESD robustness from both human body model (HBM) and machine model (MM) testing showed a U-shaped dependence as a function of sheet resistance in the range of 300–2000 Ω/sq. The physical model to explain this phenomenon was two-fold; in the first regime, a low diode series resistances and low Joule heating forming an improved diode, and in the second regime, the vertical parasitic *pnp* bipolar current gain increases above 10, leading to good vertical *pnp* gain characteristics and improved ESD results. Sub-collector sheet resistance for epitaxial sub-collectors can range from of 5 to 100 Ω/sq., dependent on the hot process and dopant type (Figure 5.16). In the majority of transistor elements, the sub-collector resistances are near the low range of values. With sub-collector resistance in the 10 Ω/sq. range, structures exhibit and ESD roll-off effect in bipolar transistors, using arsenic sub-collectors. Antimony sub-collectors have a different diffusion process, leading to a significantly higher sub-collector sheet resistance.

An experimental split was completed with arsenic and antimony sub-collector dopants with different diffusion characteristics, but the same doping concentration placed in the *p*-substrate prior to the epitaxial growth for sub-collector formation. After the hot process, the sub-collector sheet resistance is ten-fold higher for the Sb diffusion dopant implant. Experimental results show that as the size of the 'diode' (e.g. where the anode was a *p*-diffusion and the cathode was the sub-collector structure) increased in physical size, the low sheet resistance sub-collector did not maintain ESD linearity with

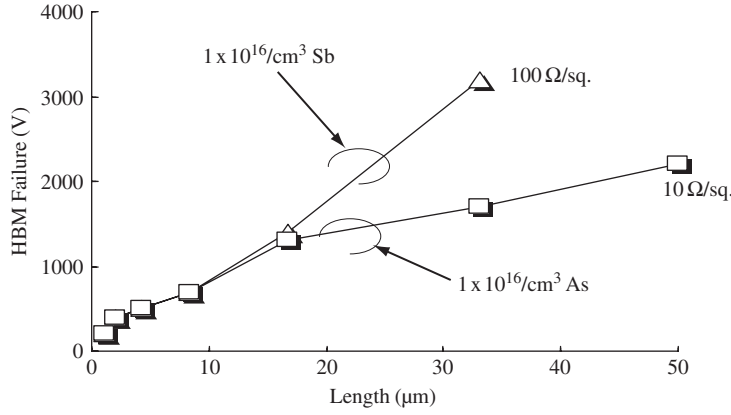

Figure 5.16 HBM ESD results as a function of different sub-collector dopants

structure size reducing the ESD robustness/length. Yet, the higher sheet resistance sub-collector dopant implant demonstrated a higher ESD/μm without sign of ESD/μm roll-off as observed in the low-resistance cathode structure [21].

In a second study (Figure 15.17), the sub-collector doping varied doping concentration by a factor of 10× per split. Experimental results in bipolar-based varactor structures showed ESD/μm nonlinear roll-off, whereas the low-dose sub-collectors remain linear. In the low sheet resistance split, HBM ESD results remained linear from 0 to 15 μm. As the length increased from 15 to 50 μm, HBM results increased to 2200 V. For the case of the higher sheet resistance sub-collector (e.g. 100 Ω/sq.), the ESD/μm increase linearly and achieve 3500 V HBM. As the implant dose was lowered by a factor of 10, the ESD results increased to 4200 V (two-fold improvement). Based on prior work of the author, this is

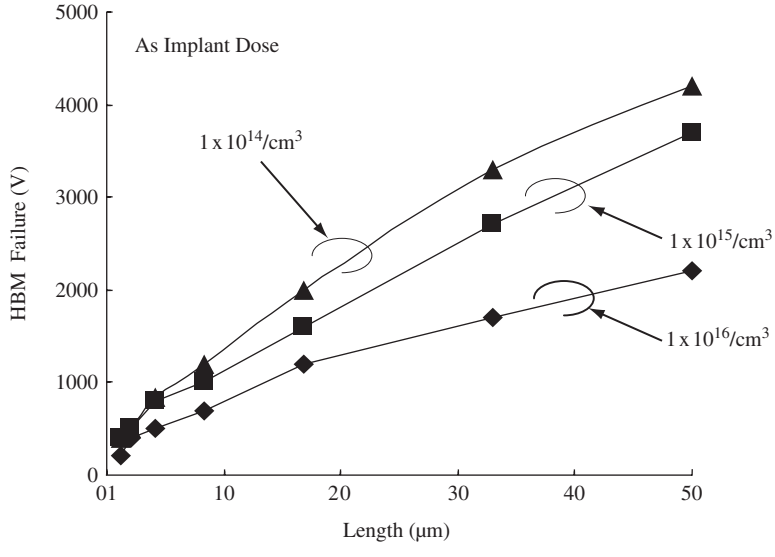

Figure 5.17 HBM ESD results as a function of different implant doses

counter-intuitive assuming the relationship of ESD/μm as demonstrated in the 200–2000 Ω/sq. sheet resistance regime. Hence, this implies that at ultra-low sheet resistances, no lateral ballasting occurs in the cathode structure. When the sheet resistance exceeds this regime, inherent lateral ballasting in the sub-collector structure leads to improved uniformity.

From this work, it can be understood that, whereas the prior measurements of retrograde and diffused *n*-wells demonstrated a U-shaped dependence as a function of the sheet resistance, these results for sub-collectors demonstrate that the ESD robustness can be lower for the lowest sheet resistance for wide ESD structures. This effect can best be explained as a nonuniform current distribution in the diode structures. As the resistance increases from 10 to 100 Ω/sq., the lateral resistance within the sub-collector introduces a ballasting effect. As the lateral resistance increases, there is less lateral current flow along the sub-collector structure. This manifests itself as the linearization with structure width [21].

A physical incremental model which explains the lateral resistor ballasting effect can be represented as a diode and resistor network (Figure 5.18). Given an anode electrical connection separated into *n* elements, the metallurgical junction can be represented as a set of parallel ideal diode elements where each *j*th element is at temperature T_j and ideal characteristics $I_j(T_j) = \mathrm{I}_0 \exp\{Vbe_j/kT_j\}$. The collector structure can be represented as a lateral resistor element between the *j*th and the $(j-1)$th diode region and between the *i*th and the $(j+1)$th diode region. This is followed by a collector series resistance which represents the series resistance. The n^+ reach-though or n^+ contact region can be represented as a lateral network or as an electrical short. As the local temperature increases along the diode stripe, the current through the segment decreases. The nonuniformity in temperature leads to a lateral current flow. The lateral resistor element $R_{\mathrm{LAT}\ j-1}(T_j)$ and element $R_{\mathrm{LAT}j+1}(T_j)$ redistribute the current from the *j*th diode element [21].

In another study, with implanted sub-collectors, similar results were achieved with elimination of the ESD roll-off effect observed in the high sub-collector dose structure (Figure 5.19). From these experimental results, sub-collector doping sheet resistance, and profile can play a significant role in the ESD robustness of a bipolar transistor, a diode-configured bipolar transistor, or even STI-bound diode elements placed in sub-collectors.

In Chapter 6, the influence of isolation structures on ESD protection will be discussed. The isolation structures include LOCOS isolation, shallow trench isolation and

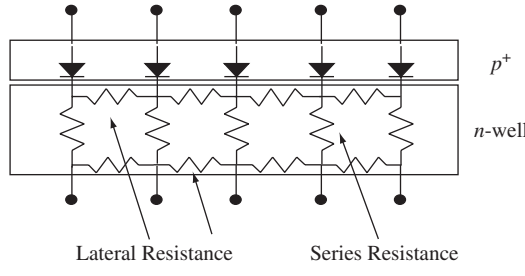

Figure 5.18 Sub-collector lateral ballasting network

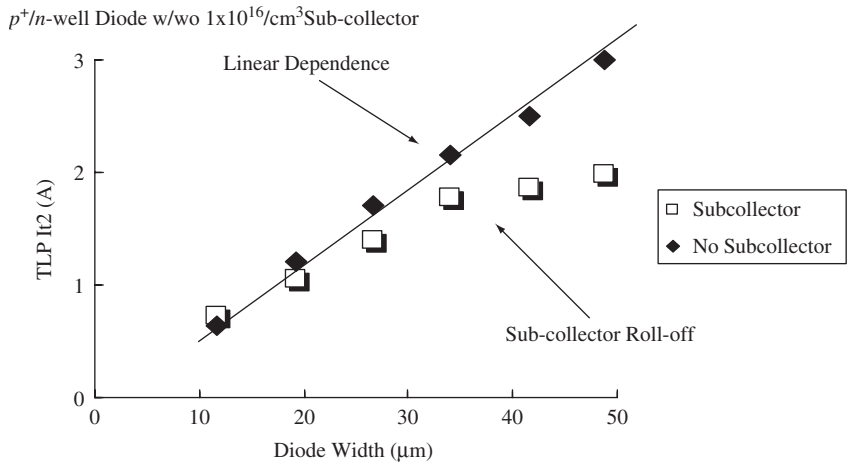

Figure 5.19 STI-bound p^+/n-well diode with/without sub-collector

deep trench isolation. These structures have significant impact on the operation of ESD networks and latchup. Distinctions are made between how ESD devices respond to LOCOS and STI. Simulation, and experimental results of LOCOS and STI structures will be discussed.

PROBLEMS

5.1. Using a Wunsch–Bell equation, derive ESD robustness scaling according to constant electric field scaling parameter α, assuming that dimensional similitude is preserved in three physical dimensions.

5.2. Using the well scaling relationship, assume that scaling of an element is scaled in only the well depth. How does the ESD robustness scale?

5.3. Using the well scaling relationship, what must the sheet resistance and film thickness be in order to maintain constant ESD robustness?

5.4. Using the well scaling relationship, with MOSFET constant electric field scaling parameter α, and defining a new well scaling parameter γ, what will the relationship of the parameters be in order to preserve the ESD robustness with one-dimensional geometric parameter scaling? Two-dimensional? Three-dimensional?

5.5. Given an n-well ballast resistor element of width W, what doping concentration should be maintained as the width is scaled as $W' = W/\alpha$ to maintain the same n-well ballast saturation current?

5.6. Given a heavily doped sub-collector with a nonlinear rolloff width effect of

$$V_{\mathrm{HBM}} = AW - BW^2$$

Derive the relationship of a width W′ for a linear element compared with W to achieve the same HBM ESD level.

REFERENCES

1. S. Dabral, R. Aslett and T. Maloney. Designing on-chip power supply coupling diodes for ESD protection and noise Immunity. *Proceedings of the EOS/ESD Symposium*, 1993; 239–249.
2. S. Dabral, R. Aslett and T. Maloney. Core clamps for low voltage technologies. Proceedings of the EOS/ESD Symposium, 1994; 141–149.
3. P. Cottrell, S. Warley, S. Voldman, W. Leipold and C. Long. N-well Design for Trench DRAM Arrays. *International Electron Device Meeting (IEDM) Technical Digest*, December 1988; 584–587.
4. S. Voldman *et al*. Retrograde well and epitaxial thickness optimization for shallow- and deep-trench collar merged isolation and node trench (MINT) SPT cell and CMOS logic technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1992; 811–815.
5. S. Voldman *et al*. Shallow trench isolation (STI) double-diode electrostatic discharge (ESD) circuit and interaction with DRAM circuitry. *Proceedings of the EOS/ESD Symposium*, 1992; 277–288; and *Journal of Electrostatics* 1993; **31**, (2–3): 237–265.
6. S. Voldman. ESD protection in a mixed voltage interface and multi-rail disconnected power grid environment in 0.5 and 0.25 μm channel length CMOS technologies. *Proceedings of the EOS/ESD Symposium*, 1994; 125–134.
7. S. Voldman. Retrograde well implants and ESD. *SEMATECH Vertical Modulated Well PTAB, SEMATECH, Austin, Texas*, November 1994.
8. S. Voldman and V. Gross. Scaling, optimization, and design considerations of electrostatic discharge protection circuits in CMOS technology. *Proceedings of the EOS/ESD Symposium*, 1993; and *Journal of Electrostatics* 1994; **33** (3) 327–357.
9. S. Voldman. Optimization of MeV retrograde wells for advanced logic and microprocessor/powerPC and electrostatic discharge (ESD). *Smart and Economic Device and Process Designs for ULSI Using MeV Implant Technology Seminar: SEMICON West, SEMICON West GENUS Seminar*, San Francisco, 1994.
10. S. Voldman. Electrostatic discharge protection, scaling, and ion implantation in advanced semiconductor technologies. *Process Integration Issues/Technical Trends Session; Proceedings of the Ion Implantation Conference (I^2CON)*, Napa, California, 1999.
11. S. Voldman, S. Furkay and J. Slinkman. Three dimensional transient electrothermal simulation of electrostatic discharge protection networks. *Proceedings of the EOS/ESD Symposium*, 1994; 246–257.
12. S. Voldman and G. Gerosa. Mixed voltage interface ESD protection circuits for advanced microprocessors in shallow trench and LOCOS isolation CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, December 1994; 811–815.
13. S. Voldman, G. Gerosa, V. Gross, N. Dickson, S. Furkay and J. Slinkman. Analysis of snubber clamped diode string mixed voltage interface ESD protection networks for advanced microprocessors. *Proceedings of the EOS/ESD Symposium*, 1995; and *Journal of Electrostatics* 1996; **38** (1–2): 3–32.
14. S. Voldman. MeV implants boost device design. *IEEE Circuits and Devices* 1995; **11** (6): 8–16.
15. S. Voldman. The impact of MOSFET technology evolution and scaling on electrostatic discharge protection. *Microelectronics Reliability* 1998; **38**: 1649–1668.
16. S. Voldman. The state of the art of electrostatic discharge protection: physics, technology, circuits, designs, simulation and scaling. *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) Symposium*, 27–29 September 1998; 19–31.
17. V. Puvvuda, V. Srinivasan and V. Gupta. A scalable analytical model for ESD *n*-well resistors. *Proceedings of the EOS/ESD Symposium*, 2000; 437–446.
18. S. Voldman *et al*. Electrostatic discharge and high current pulse characterization of epitaxial base silicon germanium heterojunction bipolar transistors. *Proceedings of the international Reliability Physics Symposium*, March 2000; 310–316.

19. S. Voldman, N. Schmidt, R. Johnson., L. Lanzerotti, A. Joseph, C. Brennan, J. Dunn, D. Harame, P. Juliano, E. Rosenbaum and B. Meyerson. Electrostatic discharge characterization of epitaxial base silicon germanium heterojunction bipolar transistors. *Proceedings of the EOS/ESD Symposium*, 2000; 239–251.
20. S. Voldman *et al*. ESD robustness of a silicon germanium BiCMOS technology. *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) Symposium*, 2000; 214–217.
21. S. Voldman, L. Lanzerotti, B. Ronan, S. St Onge and J. Dunn, The influence of process and design of sub-collectors on the ESD robustness of ESD structures and silicon germanium heterojuncion bipolar transistors in a BiCMOS SiGe technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 347–356.

6 Isolation Structures and ESD

Isolation structures influence the ESD robustness of semiconductor components from both electrical and thermal device physics. In this chapter, we will discuss the influences of isolation structures on ESD device operation. The focus of the chapter will be the influence of LOCOS isolation, shallow trench isolation, and deep trench isolation. The focus of the chapter will be to discuss the distinctions between LOCOS and STI issues, and how they influence the ESD robustness of structures.

6.1 ISOLATION STRUCTURES

Isolation structures have a significant influence on the electrical and thermal properties of a semiconductor device under high-current conditions. The geometric dimensions as well as the material properties play a role on how the structure impacts the electrical characteristics and the thermal characteristics of the semiconductor device. Isolation structures can have a significant role in the operation of a parasitic device and its impact on semiconductor devices, structures and circuits. Isolation structures can be the source of failure mechanisms in circuits where unintended or unanticipated mechanisms occur. In ESD networks, these elements can be used as devices for ESD protection. Understanding these structures and the role that isolation plays is critical in the understanding on how to obtain good ESD protection levels in a semiconductor chip. These parasitic structures can consist of guard ring structures which can be utilized for ESD protection. Isolation structures can influence latchup immunity.

Isolation structures consist of processes known as recessed oxide (ROX), local oxidation (LOCOS), and shallow trench isolation (STI). These isolation processes are typically utilized at the semiconductor surface. Additional processes exist near the surface, known as dual depth shallow trench isolation (dual depth STI), trench isolation (TI) and groove isolation structures. Additional structures used for bipolar memory, and DRAM development utilized a 'deep trench' structure which extending significantly into the semiconductor substrate. Today DT structures are utilized in DRAMs, and advanced bipolar development. This section will discuss the implications of the isolation structures on ESD and latchup.

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

6.1.1 LOCOS Isolation

Local oxidation (LOCOS) isolation (Figure 6.1) was developed to define the diffusions in MOSFET technology for transistors structures. LOCOS isolation is nonplanar, with the isolation partially formed above the semiconductor substrate surface and partially formed below. A oxide and nitride mask is formed to define the region where the isolation is to be formed. Using an oxidation process, the silicon is oxidized, leading to growth above and below the silicon surface. The isolation structure expands

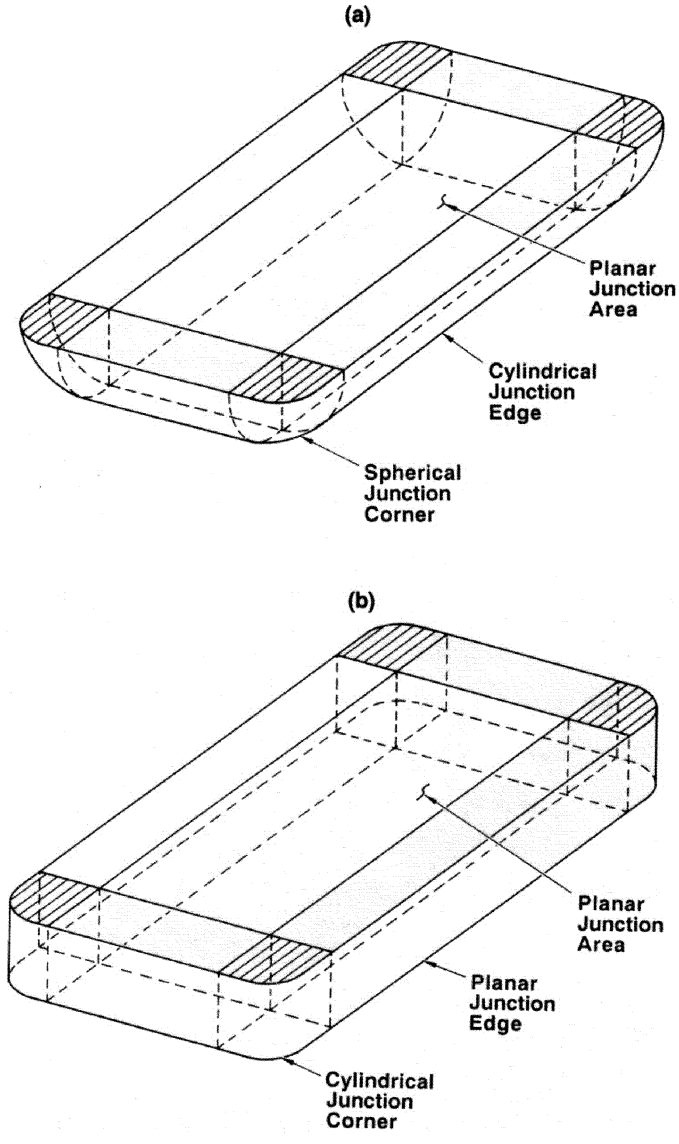

Figure 6.1 LOCOS isolation compared with shallow trench isolation

laterally, lifting the mask structure on the edge, forming a region known as the LOCOS bird's beak. The LOCOS bird's beak varies with the oxidation process. The LOCOS bird's beak was a limitation in the control of the channel width of a MOSFET transistor. This introduced variation in the channel width from the printed images, leading to variation in the channel width ΔW and device-to-device matching variation Δ (ΔW).

The LOCOS bird's beak impacted ESD protection by producing variation between the junction edge and the salicide edge. LOCOS isolation lead to ESD failure mechanism due to the relative distance between the salicide (self-aligned silicide) and the metallurgical junction. In STI structures, the sidewall of the isolation is vertical, leading to small variations in the location of salicide formation and definition of the junction edge. STI-defined junctions have eliminated the issue of the lateral location of the salicide and the metallurgical junction edge by having a well-controlled lateral variation.

As junctions are scaled in depth in a LOCOS process, the curvature on the edge of the metallurgical junction leads to higher electric fields with scaling [1,2]. The electric field enhancement at the curvature region can lead to electrical breakdown, hot electron effects, and nonuniform current distribution. In two dimensions, the junction is enhanced on the edge, and at the corners, there is a further enhancement of the electrical field. From Gauss' law

$$\nabla \bullet E = \frac{\rho}{\varepsilon}$$

Addressing the junction curvature at the edges, the divergence can be expressed in cylindrical coordinates, assuming only a radial component

$$\frac{1}{r}\frac{\partial}{\partial r}[rE(r)] = \frac{\rho(r)}{\varepsilon}$$

Solving for the electric field

$$\frac{\partial}{\partial r}[rE(r)] = r\frac{\rho(r)}{\varepsilon}$$

Integrating both sides

$$\int_{r_j}^{r} \mathrm{d}r \frac{\partial}{\partial r}[rE(r)] = \int_{r_j}^{r} r\frac{\rho(r)}{\varepsilon}\mathrm{d}r$$

where the solution can be expressed as a function of the radius of curvature of the junction r_j, and a constant of integration K

$$E(r) = \frac{1}{r}\int_{r_j}^{r} r\frac{\rho(r)}{\varepsilon}\mathrm{d}r + \frac{K}{r}$$

S. K. Ghandi [3] showed that this can be expressed for a one-sided junction, the cylindrical normalized breakdown voltage can be expressed as

$$\frac{V_{\text{cyl}}}{V_{\text{B}}} = \frac{1}{2}\left(\eta^2 + 2\eta^{6/7}\right)\ln\left(1 + \frac{2}{\eta^{8/7}}\right) - \eta^{6/7}$$

where the breakdown voltage for a one-sided junction is expressed as

$$V_{\text{B}} = \frac{E_{\text{m}} W}{2} = \frac{\varepsilon {E_{\text{m}}}^2}{2qN_{\text{a}}}$$

and the junction curvature ratio factor

$$\eta = \frac{r_j}{W_m}$$

At the corners of the junction, spherical geometry can be assumed where Gauss' law can be expressed in spherical coordinates as

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 E(r)\right] = \frac{\rho(r)}{\varepsilon}$$

where, solving for the electric field by integration, the solution can be put in the form, where r_j is the radius of curvature of the junction

$$E(r) = \frac{1}{r^2}\int_{r_j}^{r} r^2 \frac{\rho(r)}{\varepsilon}\mathrm{d}r + \frac{K}{r^2}$$

with the spherical expression for normalized breakdown voltage

$$\frac{V_{\text{sph}}}{V_{\text{B}}} = \eta^2 + 2.14\eta^{6/7} - \left(\eta^3 + 3\eta^{13/7}\right)^{2/3}$$

As technologies scale, the radius of curvature decreases, lowering the breakdown voltage of both the cylindrical edges and spherical corners. For a reverse-biased structure, this leads to initiation of ESD failures at lower voltage conditions. In diode structures, the lowering of the reverse-bias breakdown voltage can lead to ESD failures at lower voltages. In MOSFETs, the source/drain junction depth is scaled with MOSFET gate dielectric thickness, and channel length. As a result, the MOSFET snapback voltage decreases, leading to lower MOSFET avalanche breakdown. This scaling effect can be eliminated in diode structures by avoidance of the cylindrical

and spherical electric field enhancements by extension of the isolation below the metallurgical junction.

6.1.2 LOCOS-bound ESD Structures

A significant number of structures are suitable for ESD protection and are of interest for understanding of failure mechanisms. LOCOS-defined ESD structures can consist of thick oxide lateral *pnp* and *npn* bipolar transistors, p^+/*n*-well diodes, n^+/substrate diodes, and LOCOS-defined MOSFET structures. LOCOS-defined structures can include structures that extend well below the LOCOS and junction depth; *n*-well-to-*n*-well and *n*-well-to-LOCOS-defined n^+ junction lateral bipolar transistors also serve for ESD protection in a LOCOS-defined technology.

6.1.2.1 p^+/n-well junction diodes

LOCOS-bound p^+/*n*-well junction diodes are formed by using the LOCOS isolation to define the anode region as well as the cathode region of the diode structure [4]. In a forward-bias operation mode, majority-carrier holes are injected from the anode to the cathode across the metallurgical junction (Figure 6.2). Minority-carrier holes diffuse vertically and laterally in the *n*-well cathode region. Holes that reach the STI-bound n^+ contact recombine. Holes that diffuse vertically recombine in the *n*-well region or diffuse to the *n*-well-to-substrate junction. The ratio of lateral and vertical hole current density is a function of the *n*-well profile, sheet resistance and relative distance between the p^+ and n^+ contact diffusion.

The ESD robustness of the LOCOS-bound p^+/*n*-well diode is a function of the p^+ diode length, width, finger number, p^+ to *n*-well contact spacing, single-sided or wrap-around configuration, and the sheet resistance of the *n*-well. A p^+/*n*-well diode structure performs best as a device when the n^+ well contact wraps around the p^+ diffusion, allowing for uniform current distribution. ESD robustness of the p^+/*n*-well diode is a weak function of the p^+-to-*n*-well contact space (e.g. the LOCOS width). The

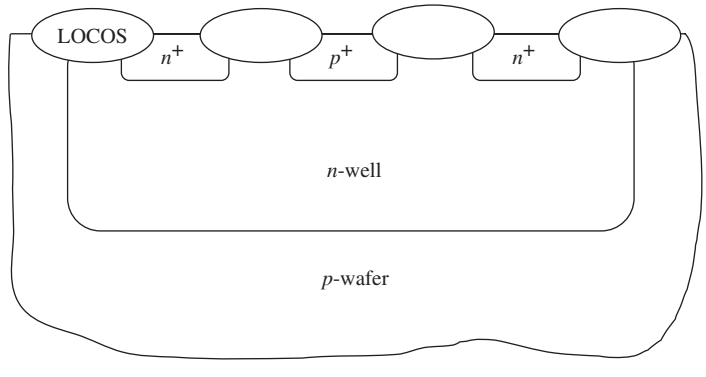

Figure 6.2 p^+/*n*-well diode structure

ESD failure level is linear with diode area for approximately the equivalent perimeter. Experimental results also show that as the p^+/n-well diode perimeter increases, the ESD robustness increases significantly. Additionally, for a fixed area, as the ratio of perimeter to area increases, both HBM and MM experimental results showed significant improvement.

In LOCOS-defined p^+/n-well diodes, to minimize the impact of the silicide penetration to the metallurgical junction near the LOCOS bird's beak edge, masks can be used at the perimeter of the diffusion to prevent failure mechanisms. Since the perimeter of the LOCOS-defined diffusion has the highest electric field, using a silicide block mask near the LOCOS bird's beak edge can allow for the removal of the silicide near the diode perimeter. The width of the mask can allow for increasing the physical lateral distance between the silicide and the edge of the metallurgical junction.

Figure 6.3 compares the ESD robustness of a diode structure in LOCOS and shallow trench isolation for different features sizes. The LOCOS diode ESD performance exceeded the shallow trench isolation performance for the same structure size.

6.1.2.2 n^+ junction diodes

LOCOS-bound n^+ junction/substrate diodes provide good characteristics for injection of current into a p^- or p^-/p^+ substrate. In LOCOS technology, the usage of an n-junction for ESD protection is a concern because of the variation of the salicide relative to the metallurgical junction edge. With the introduction of salicide, n^+ diffusions provide good ESD robustness for negative discharge events. As the reverse bias increases across the n^+ diode, the depletion width of the metallurgical junction extends into the n^+-doped region. With the LOCOS bird's beak variation, and n^+ implant lateral diffusion variation, the lateral distance between the silicide and the metallurgical

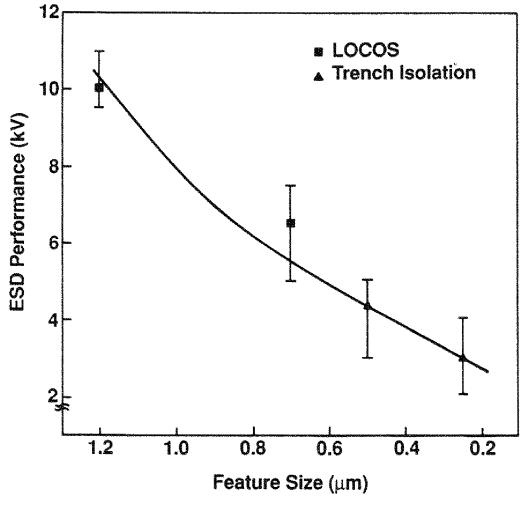

Figure 6.3 ESD performance and feature size (LOCOS–STI comparison)

junction can vary, increasing the risk of ESD failure mechanisms. As in the LOCOS-defined p^+/n-well diode structure, a mask to prevent silicide formation near the edge of the diode can minimize ESD failures.

6.1.2.3 n-well/substrate diodes

In LOCOS technology, n-well diodes in p^+ substrates, were used to avoid the penetration of the aluminum metallurgy and silicide into the metallurgical junction of an n^+ diffusion. In LOCOS-defined diode structures with tapered via formed in aluminum metallurgy, vertical penetration from aluminum spiking is possible in reverse bias when the temperature of the via structure exceeds the melting temperature. This is a larger concern in n^+ junctions in reverse-bias HBM events, or in CDM events. To avoid this concern, early semiconductor technologies utilized n-well diodes instead of n^+ junction diodes.

N-Well-to-substrate metallurgical junctions were not adjacent to the LOCOS edge, and were significantly deeper which avoided this concern [4]. In a LOCOS-defined technology, the n-well extends below the LOCOS isolation and n^+ diffusion depth. The LOCOS-defined n^+ diffusion serves as the contact to the n-well diode structure. Because the n-well structure is significantly wider than the n^+ diffusion and lateral n-well out-diffusion the spacing between the silicide edge and the n-well-to-substrate metallurgical junction is significantly spaced to avoid silicide induced ESD failure mechanisms. Because the n-well dopant extends below the LOCOS isolation region, the n-well diode structures have junction curvature at the diode perimeter at the bottom of the implant. The advantage of the n-well structure is that the heat is dissipated well below the LOCOS structure, and provides a good junction area for injection into the substrate. LOCOS-defined well structures provide improved injection characteristics compared to the n^+ diffusion – to-substrate diode structure. A disadvantage of the n-well structure is there exists a minimum scalable length preventing scaling of the length of the n-well diode, and typically a significant higher perimeter capacitance compared to the LOCOS-bound diode structure. This can be compensated by using a shorter width.

6.1.2.4 Lateral n-well-to-n-well bipolar ESD element

LOCOS-bound lateral n-well-to-n-well bipolar ESD elements are effective solutions to provide ESD protection for negative discharge events [4]. The n-well extends below the LOOS region allowing for electrons current to flow between adjacent n-well structures without impediment of the LOCOS structure. A lateral npn structure consists of a first n-well serving as an emitter, a p-substrate serving as the base, and a second n-well serving as a collector. Electrons flow laterally from the n-well emitter to the n-well collector structure. The n-well emitter can be an LOCOS-defined n-well connected to an input pad, and the n-well collector can be any adjacent n-well connected to a power supply or power rail, or a n-well guard ring structure. In this case, the electrons must flow at the LOCOS-bound n-diffusion diode-to-substrate metallurgical junction to the adjacent n-well region. In this structure, the emitter area is the LOCOS-defined n^+ junction area.

6.1.2.5 Lateral pnp bipolar ESD element

LOCOS-bound lateral *pnp* transistors are formed using the *p*-channel MOSFET source/drain implant in an *n*-well region. A p^+ diffusion serves as the emitter, and a second p^+ diffusion serves as the collector, while the *n*-well serves as the base region. The lateral *pnp* transistor has a lateral component whose emitter area is the product of the diffusion depth and width.

6.1.2.6 Thick oxide MOSFET ESD element

LOCOS-defined thick oxide MOSFET transistors (Figure 6.4) consists of a first and second *n*-diffusion separated by a LOCOS isolation region. A metal or polysilicon gate structure is placed on the LOCOS isolation forming a thick oxide gate structure. The LOCOS-defined thick oxide transistor has a high MOSFET threshold voltage due to the LOCOS gate dielectric region. These devices are suitable for a primary stage of an NMOS ESD network [5–12]. The thick oxide MOSFET was used as ESD networks in the 3–1 μm technology generations. The device operates as a lateral bipolar transistor. These devices were vary sensitive to the LOCOS bird's beak, silicide formation, and the spacing from the drain contact to the diffusion edge. The physics of the drain contact, the salicide, the edge of the isolation and the metallurgical junction influenced the experimental results, which gave conflicting results. The three major design variables for these structures were the length, the width, and the drain contact-to-diffusion edge spacing. Weston *et al.* showed that as the base width of the lateral device (thick oxide MOSFET) decreased, trigger voltage decreased in a linear fashion. At very short dimensions, leakage was evident [12]. At very long dimensions, gate oxide breakdown occurred. From the experimental results on LOCOS-defined thick oxide MOSFET devices, the operation is very dependent on the control of the spacing between the salicide edge and the metallurgical junction, which is modulated by the LOCOS bird's beak. Improvement in this structure was achieved using silicide masking, silicide removal and abrupt junctions [5–12].

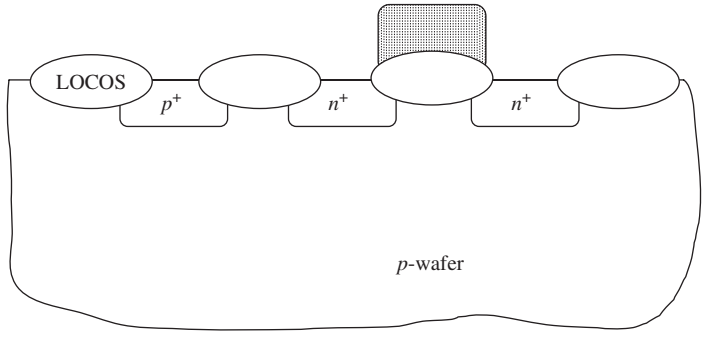

Figure 6.4 Thick oxide MOSFET structure

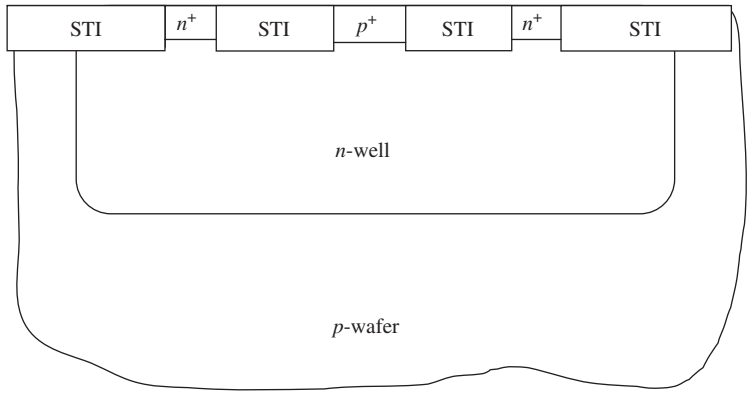

Figure 6.5 STI-bound p^+/n-well diode

6.2 SHALLOW TRENCH ISOLATION

Shallow trench isolation (STI) was implemented into semiconductors for MOSFET constant electric field scaling and junction scaling as well as to address issues connected with isolation processes such as LOCOS isolation [13]. This technology (Figure 6.5) was first initiated in the late 1980s and integrated into CMOS DRAM and logic technology in the early 1990s [13, 14]. The construction of the first STI-defined ESD networks were initiated by Hargrove and Voldman [15,16].

There are a number of features of STI structures which are significantly different from LOCOS isolation. Where it was assumed ESD networks would be easily mapped from LOCOS-defined technology to STI-defined technology, significant ESD understanding was required to optimize the ESD networks in STI-defined technology [15–24]. LOCOS isolation was nonplanar with the isolation partially formed above the semiconductor substrate surface and partially formed below. The topography of the LOCOS hampered its ability for scaling in future technology generations. STI structures are planar with the semiconductor device surface. A second feature known as LOCOS bird's beak was a limitation in the control of the channel width of a MOSFET transistor. This introduced variation in the channel width from the printed images, leading to variation in the channel width ΔW and device-to-device matching variation Δ (ΔW). STI structures could maintain planarity with different widths between adjacent structures.

The LOCOS bird's beak impacted ESD protection by producing variation between the junction edge and the salicide edge [5–12]. LOCOS isolation lead to ESD failure mechanism due to the relative distance between the salicide and the metallurgical junction. In STI structures, the sidewall of the isolation is vertical leading to small variations in the location of salicide formation and definition of the junction edge. STI-defined junctions have eliminated the issue of the lateral location of the salicide and the metallurgical junction edge by having a well-controlled lateral variation.

Another feature of STI structures and STI-defined junctions is that implants formed at the device surface, such as p^+ and n^+ implants, have metallurgical junctions which do not extend below the STI structure. The STI depth is a function of the aspect ratio of width to depth that is acceptable in a manufacturing environment.

With the STI depth below the metallurgical junction, MOSFET and other surface implants are scalable without electric field enhancements at the junction edge. As junction are scaled in depth in a LOCOS process, the curvature on the edge of the metallurgical junction leads to higher electric fields with scaling. The electric field enhancement at the curvature region can lead to electrical breakdown, hot electron effects, and nonuniform current distribution. In two dimensions, the junction is enhanced on the edge, and at the corners, there is a further enhancement of the electrical field. Additionally, when the junctions extend below the isolation structure, the perimeter capacitance exceeds the area capacitance. In STI junctions, there is no radius of curvature formed at the metallurgical junction–STI interface which could lead to electrical breakdown (except the cylindrical curvature at the corners). Additionally, the perimeter capacitance is significantly lower (or negligible) compared with the area capacitance term.

6.2.1 Shallow Trench Isolation Pull-down

Shallow trench isolation avoided the lateral bird's beak effect of LOCOS isolation structures, but introduced a new issue [24]. In the formation of shallow trench isolation, the objective is to achieve a planar interface between the insulator regions and the silicon areas. Etch and masking processes, and salicide formation volumetric changes, however, cause a nonplanar intersection of the isolation and silicon areas. In early development of shallow trench isolation technology, a pad nitride mask is formed over the silicon areas to avoid silicon etching; after oxide isolation filling, this mask is removed. The STI regions is exposed during the etch process, leading to nonplanar STI edges where the silicon region extends above the isolation edges.

In the shallow trench isolation – silicon diffusion regions, the conformally deposited refractory metal to form a silicide is deposited over the oxide regions and the STI-diffusion interface. In the formation of the silicide, the silicon laterally diffuses to react with the refractory metal leading to interface recession. These effects can lead to STI-perimeter diode leakage (Figure 6.6).

In MOSFET gate structures, gate oxide deposition leads to MOSFET 'gate wrap-around' on the edges. In the ungated edges, as in the diffusion regions, the conformally

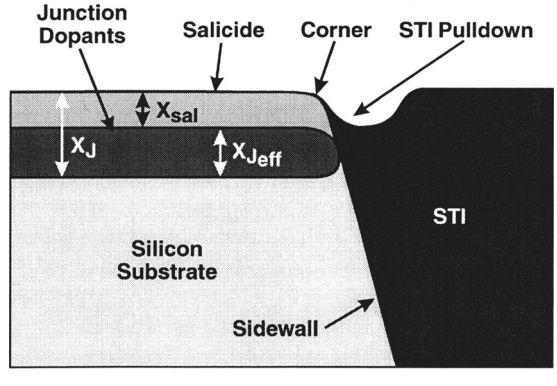

Figure 6.6 Shallow trench isolation and STI pull-down

deposited refractory metal to form the silicide is deposited over the oxide regions and the STI-diffusion interface. In the gated regions, the MOSFET gate structure wraps around the edge of the isolation on the channel sidewalls, leading to an enhanced electric field at the edge of the MOSFET structure. This can lead to higher current density at the edges of the MOSFET, as well as a three-dimensional MOSFET gate-induced drain leakage [25].

STI pull-down can be minimized with proper selection of the etch and planarization process. The leakage effect and nonideal electrical characteristics can also be minimized by placement of the metallurgical junction of significant distance from the silicide formation depth.

6.2.2 STI-bound ESD Structures

A significant number of structures are suitable for ESD protection and are of interest for understanding of failure mechanisms. STI-defined structures can consist of thick oxide lateral *pnp* and *npn* bipolar transistors, p^+/*n*-well diodes, n^+/substrate diodes, and STI-defined MOSFET structures. STI-defined structures can include structures that extend below the STI depth, *n*-well-to-*n*-well and, *n*-well-to-STI-defined n^+ junction lateral bipolar transistors also serve for ESD protection in a STI-defined technology.

In STI-bound ESD structures, the STI-pull down effect can lead to a lowering of the ESD robustness of the structure [24]. Evidence of the STI pull-down impact to the ESD structures will be evident from nonideal diode *I–V* characteristics, and increased diode leakage current. If the STI pull-down effect is minimized, STI-bound diodes will typically not show perimeter failure mechanisms as observed in LOCOS-defined ESD diode junctions.

6.2.2.1 p^+/n-well junctions

With the STI structure extending beyond the junction region, lateral thermal transport is impacted by the isolation structure. The STI region is filled with an insulator structure which impacts the lateral current flow of carriers under the isolation region, as well as the thermal flux [18,20,21] (Figure 6.7a).

In Figure 6.7(b), thermal simulation of a STI isolation and an equivalent LOCOS isolation are compared for a p^+/*n*-well diode in forward bias. From the results, it can be observed that the temperature under the STI region is higher than the LOCOS-defined junction region.

In Figure 6.8, the peak temperature is plotted as a function of time for STI and LOCOS isolation. As observed in Figure 7(b), the peak temperature during an ESD event is higher for STI isolation compared to LOCOS isolation.

In Figure 6.9(a–c), emission microscope photon emissions and the Kelvin force probe microscope image of the STI-bound diode are shown. The KPFM image has removed the oxide region, leaving a view of the silicon region where electrothermal failure occurs in STI-bound p^+/*n*-well diodes. The electrothermal failure occurs from the p^+ contact of the anode to the n^+ contact of the cathode. The failure path is through the p^+ diode junction, under the STI region to the n^+ well contact.

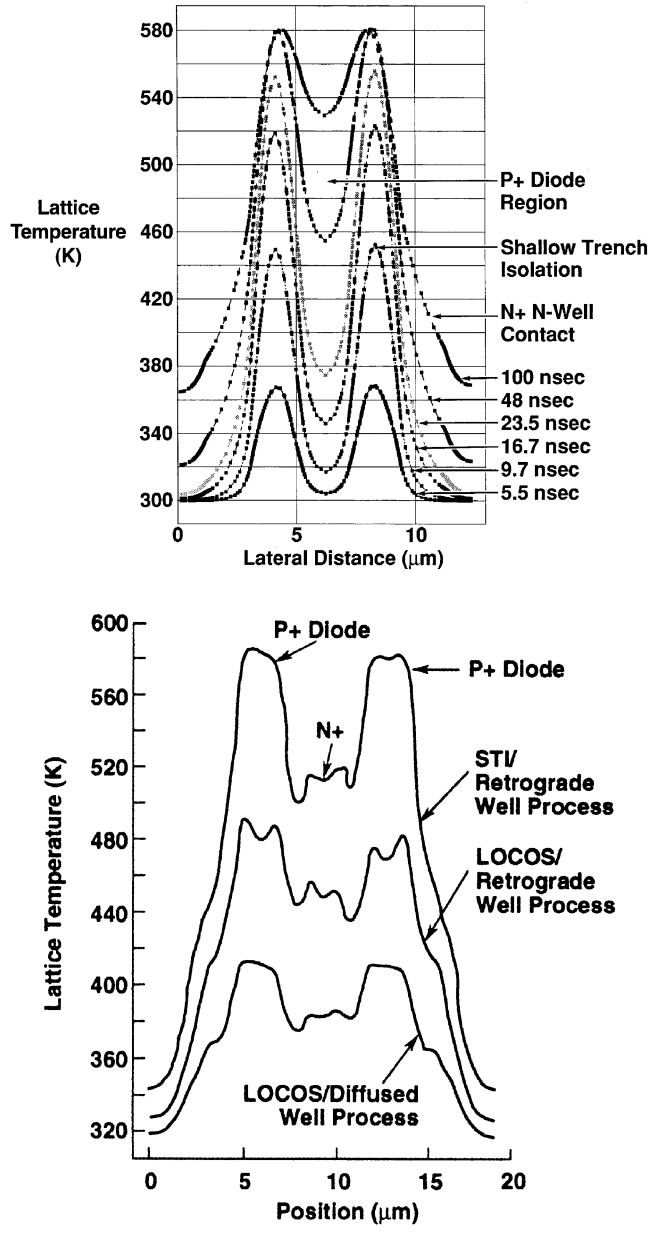

Figure 6.7 (a) Electrothermal simulation of STI p^+/n-well diode; (b) Self-heating simulation comparing LOCOS with STI for a p^+/n-well diode in a retrograde well

In an STI-bound p^+/n-well junction diode is formed by using the STI-region to define the anode region as well as the cathode region of the diode structure [15–22]. In a forward-bias operation mode, majority-carrier holes are injected from the anode to the cathode across the metallurgical junction. Minority-carrier holes diffuse vertically and

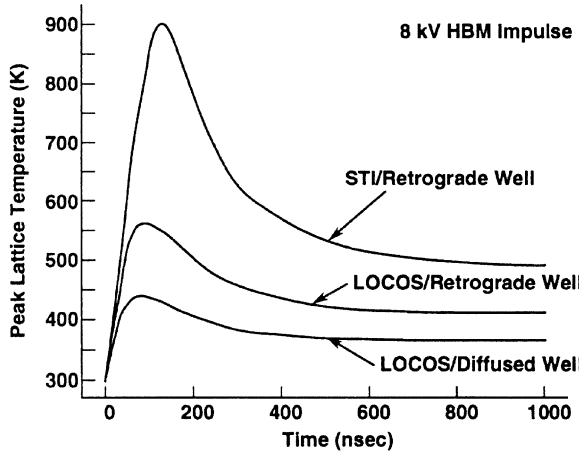

Figure 6.8 Electrothermal simulation of p^+/n-well diode comparing LOCOS and STI structure

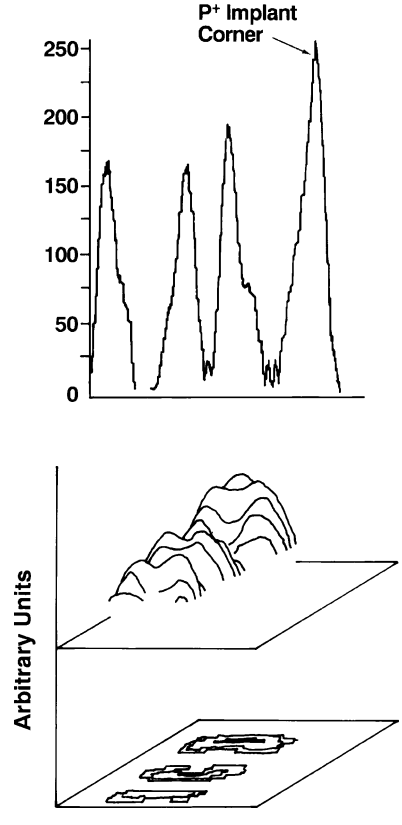

Figure 6.9 (Continued)

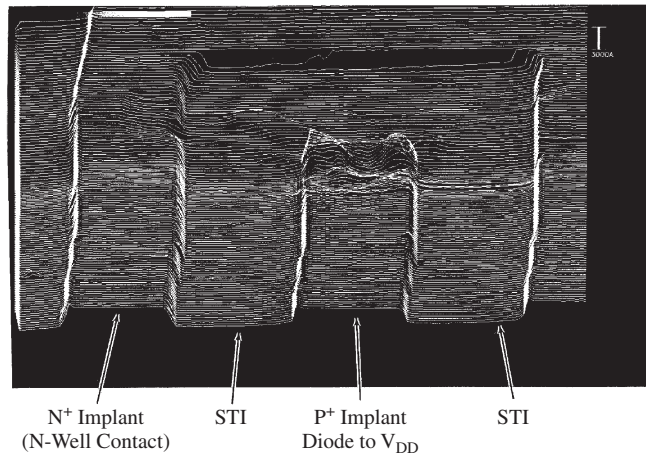

Figure 6.9 (a) Emission microscope photon emissions of a p^+/n-well diode structure; (b) emission microscope photon mapping; (c) Kelvin force probe microscope image

laterally in the n-well cathode region. Holes that reach the STI-bound n^+ contact recombine. Holes that diffuse vertically recombine in the n-well region or diffuse to the n-well-to-substrate junction. The ratio of lateral and vertical hole current density is a function of the n-well profile, sheet resistance and relative distance between the $p^+$and n^+ contact diffusion.

The ESD robustness of the STI-bound p^+/n-well diode is a function of the p^+ diode length, width, finger number, p^+-to-n-well contact spacing, single-sided or wrap-around configuration, and the sheet resistance of the n-well [15–24]. A p^+/n-well diode structure performs best as a device when the n^+ well contact wraps around the p^+ diffusion, allowing for uniform current distribution. ESD robustness of the p^+/n-well diode (Figure 6.10) is a weak function of the p^+-to-n-well contact space (e.g. the STI width). The ESD failure level is linear with diode area for approximately

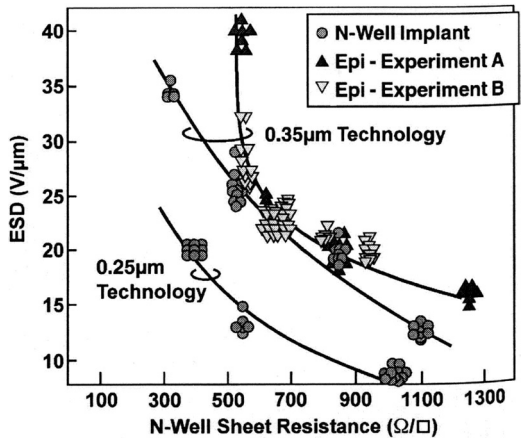

Figure 6.10 ESD and n-well sheet resistance

the equivalent perimeter. Experimental results also show that as the p^+/n-well diode perimeter increases, the ESD robustness increases significantly. Additionally, for a fixed area, as the ratio of perimeter to area increases, both HBM and MM experimental results show significant improvement. Additionally, ESD robustness of these structure increases for n-well sheet resistances in the range of 300–1000 Ω/sq. [15–21].

6.2.2.2 n^+ junction diodes

STI-bound n^+ junction/substrate diodes provide good characteristics for injection of current into a p^- or p^-/p^+ substrate [15–18]. In an STI-bound junction, there is no issue of salicide penetration on the edge of the structure. In LOCOS technology, the use of an n-junction for ESD protection was a concern because of the variation of the salicide relative to the metallurgical junction edge. With the introduction of tungsten contacts, salicide and STI-isolation, STI-bound n^+ diffusions provide good ESD robustness for negative discharge events. Tungsten (W) stud contacts have high melting temperature (e.g. 3400 °C) with no tendency for contact spiking effects. The W stud contact combined with the use of STI isolation and salicide, the effective junction depth (e.g. distance between the silicide atoms and the metallurgical junction) provides no failure mechanism at low ESD levels. Experimental results showed testing of n^+ diffusion to −10 kV HBM levels without failure in a STI-defined DRAM technology [17]. STI-bound diode structures are a function of the length, and width. These structures are not sensitive to end effects as was evident in the STI-bound p^+ diode structures.

6.2.2.3 n-well/substrate diodes

In LOCOS technology, n-well diodes in p-substrate were used to avoid the penetration of the aluminum metallurgy and silicide into metallurgical junction of an n^+ diffusion. n-Well-to-substrate metallurgical junctions were not adjacent to the LOCOS edge as well as significantly deeper which avoided this concern. In an STI-defined technology, the n-well extends below the STI-depth. The STI-defined n^+ diffusion serves as the contact to the n-well diode structure. In an STI-defined technology, the n-well-to-STI interface is well below the surface. Because the dopant extends below the STI region, the n-well diode structures (Figure 6.11) have junction curvature at the diode perimeter at the bottom of the implant, but no curvature at the STI-to-n-well interface. The advantage of the n-well structure is that the heat is dissipated below the STI structure, and provides a good junction area for injection into the substrate. A disadvantage of this structure is there exists a minimum scalable length preventing scaling of the length of the n-well diode, and typically a significant higher perimeter capacitance compared to the STI-bound diode structure. ESD robustness of the n-well-to-substrate diode in a p^-/p^+ substrate shows HBM levels over −10 kV HBM [17].

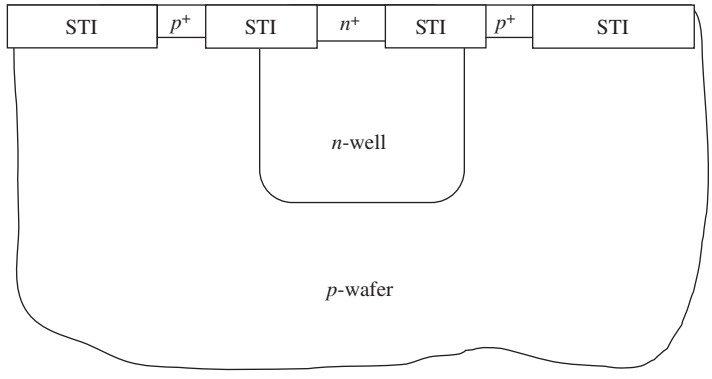

Figure 6.11 *n*-well diode structure

6.2.2.4 Lateral n-well-to-n-well bipolar ESD element

STI-bound lateral *n*-well-to *n*-well bipolar ESD elements (Figure 6.12) are effective solutions to provide ESD protection for negative discharge events. The *n*-well extends below the STI region, allowing for electrons current to flow between adjacent *n*-well structures without impediment of the STI structure. A lateral *npn* structure consists of a first *n*-well serving as an emitter, a *p*-substrate serving as the base, and a second *n*-well serving as a collector. Electrons flow laterally from the *n*-well emitter to the *n*-well collector structure. The *n*-well emitter can be an STI-defined *n*-well connected to an input pad, and the *n*-well collector can be any adjacent *n*-well connected to a power supply or power rail, or an *n*-well guard ring structure [17].

The ESD robustness of this structure is a function of the spacing between the adjacent wells (Figure 6.13). Experimental results on a p^-/p^{++} substrate wafer shows that the ESD robustness achieves HBM levels over -10 kV for spacings of 6 μm. As the effective base width is reduced from 6 to 3 μm, ESD robustness decreases [17]. This is

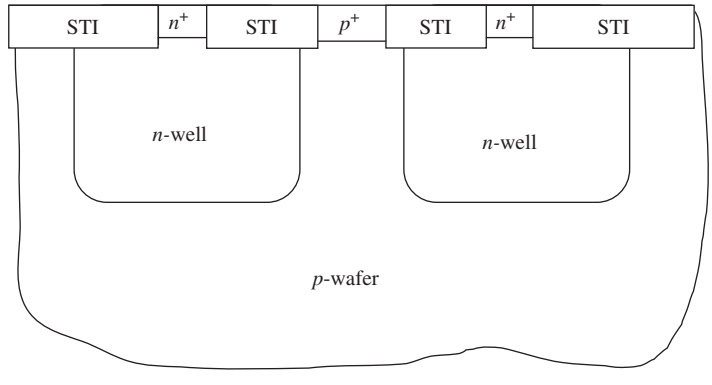

Figure 6.12 *n*-well to *n*-well diode structure

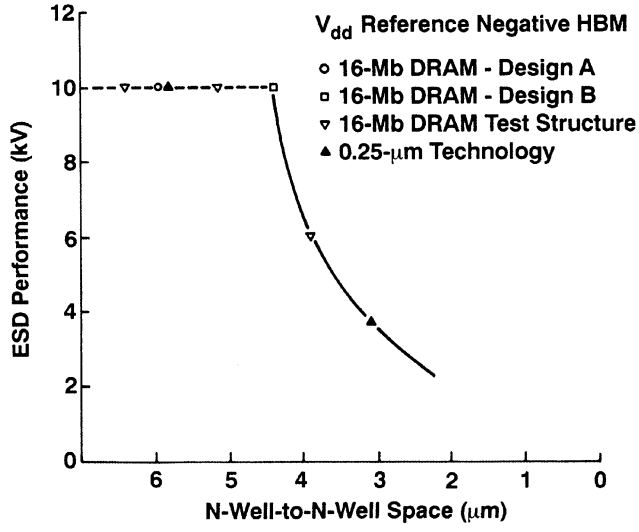

Figure 6.13 *n*-well to *n*-well ESD structure HBM negative results as a function of well-to-well spacing

also a function of the depth of the well. As the well depth is scaled, the spacing dependence is equivalent, but the ESD levels are lower [17].

In the lateral *n*-well-to-*n*-well structure, the failure analysis shows the silicon melting regions between the two *n*-wells during ESD events. This is evident from Figure 6.14.

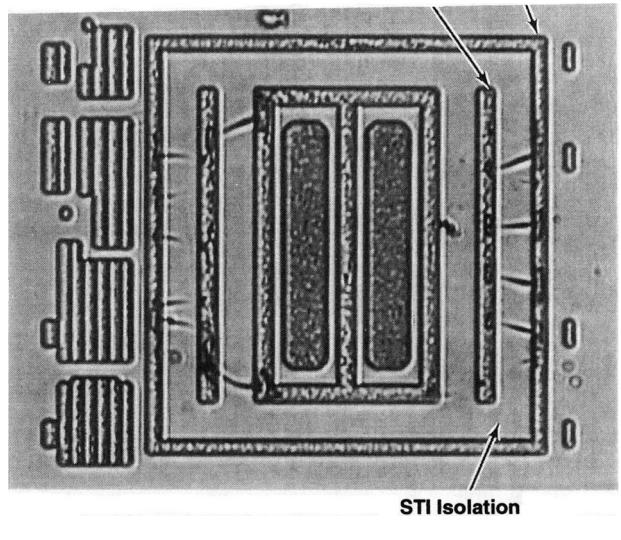

Figure 6.14 Failure analysis of ESD *n*-well-to-*n*-well structure

6.2.2.5 Lateral n^+-to n-well bipolar ESD element

STI-bound lateral *n*-diffusion-to-*n*-well bipolar ESD elements are effective solutions to provide ESD protection for negative discharge events. Using an STI-defined *n*-diffusion, electrons current flows to an adjacent *n*-well structure. A lateral *npn* structure consists of a first STI-bound *n*-diffusion serving as an emitter, a *p*-substrate serving as the base, and an *n*-well serving as a collector. Electrons flow laterally from the *n*-diffusion emitter to the *n*-well collector structure. The *n*-diffusion emitter can be connected to an input pad, and the *n*-well collector can be any adjacent *n*-well connected to a power supply or power rail, or an *n*-well guard ring structure. In this case, the electrons must flow vertically at the STI-bound *n*-diffusion diode-to-substrate metallurgical junction, and then laterally to the adjacent *n*-well region below the STI region. In this structure, the emitter area is the STI-defined n^+ junction area.

6.2.2.6 Lateral pnp bipolar ESD element

In STI technology, the depth of the isolation extends below the p^+ diffusion depth in semiconductor devices. The formation of a lateral *pnp* transistor (Figure 6.15) is formed by two closely spaced STI-defined p^+ implants in an *n*-well region. When the emitter–base junction is forward biased, the holes will diffuse into the well region. In a retrograde well, the vertical built-in field will add a drift current term opposite to the downward diffusion current term. Holes will flow laterally under the STI region to the p^+ collector region [22].

ESD robustness of ESD networks using the STI-defined lateral *pnp* device exceeded 7 kV HBM in a 0.35-μm channel length technology. Atomic force microscope mappings demonstrated the ESD failure occurs under the STI region (Figure 6.16).

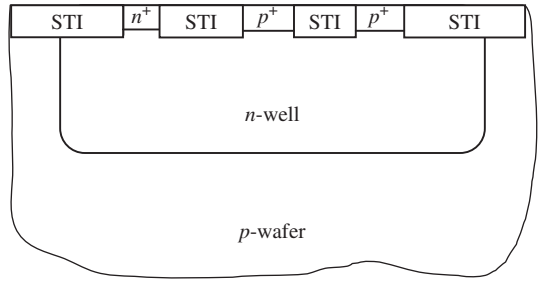

Figure 6.15 Lateral STI-defined *pnp* structure

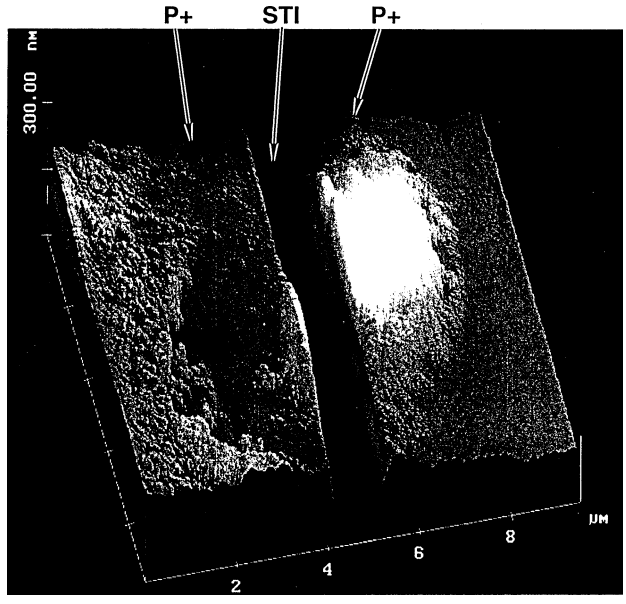

Figure 6.16 Atomic force microscope (AFM) image of STI-defined lateral *pnp* ESD structure

6.3 DEEP TRENCH ISOLATION

Deep trench (DT) isolation was implemented into semiconductors for the reduction of capacitance in bipolar technologies in the early 1980s. Bipolar memory static memory cells initially used 'diffused isolation.' Diffused isolation structures consisted of a sub-collector region, a reach-through implant and a connecting implant between the sub-collector and the reach-through implant. This 'diffused isolation' collector structure added significant parasitic collector-to-substrate capacitance, impacting the performance of bipolar memory. With the introduction of trench isolation which extended below the sub-collector, the sub-collector-to-substrate capacitance is reduced. The development of the deep trench structure transitioned with development from oxide-filled trench, polyimide-filled deep trench to polysilicon-filled deep trench technology. This technology was also developed in parallel in DRAM technology [14]. Today, deep trench structures are common practice in BiCMOS, and BiCMOS silicon–germanium technology [26–34]. A key focus in trench isolation is the reduction in the collector capacitance, and the impact on transistor heating [26–29]. Deep trench technology also influences latchup and ESD robustness [30–34].

6.3.1 Deep Trench Guard Ring Structures

Deep trench isolation can be used for improved isolation of well regions, as a means to establish isolated epitaxial regions, or as independent guard ring structures [31]. The

deep trench structures can be used to minimize lateral parasitic devices, injection into the substrate, and to improve the latchup robustness of a technology.

Guard ring structures (Figure 6.17) can be used to minimize minority-carrier injection into the substrate which can lead to noise injection or latchup. Deep trench guard rings can be used around ESD and I/O elements to reduce the latchup sensitivity and to provide improved isolation between circuits. A measure of the value of the ability of a guard ring to isolate circuits is by evaluation of the 'guard ring efficiency.'

Given a minority-carrier in a substrate, a carrier either remains within the deep trench guard ring structure or escapes past the guard ring structure. Hence the probability of capture and the probability of escape is equal to unity.

$$P(\text{capture}) + P(\text{escape}) = 1$$

In the case of a *p*-substrate, the minority carrier electron can be 'captured' either by recombination within the DT guard ring, or are collected at a metallurgical junction within the DT guard ring.

$$P(\text{capture}) = P(\text{collected}) + P(\text{recombination})$$

The probability that an electron escapes beyond the DT guard ring is the probability that an electron recombines outside the DT guard ring and the probability that an electron is collected at a structure outside the guard ring.

$$P(\text{escape}) = P(\text{collected}) + P(\text{recombination})$$

Hence the guard ring efficiency of a deep trench structure is related to the probability that an electron is trapped within the guard ring structure.

Assuming an injection source inside the DT guard ring, and a collecting structure outside the DT guard ring, an efficiency metric can be defined as F, the ratio of injected current to the collected current

$$F = \frac{I(\text{injected})}{I(\text{collected})}$$

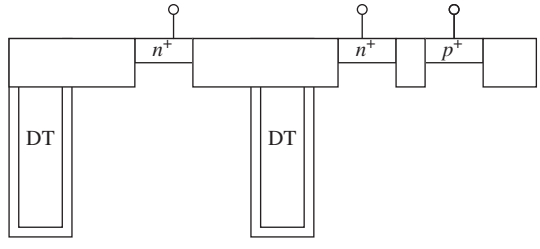

Figure 6.17 Deep trench structure guard ring

In this case, the inverse relationship $1/F$ is the amount of current collected outside the DT guard ring at an outer structure, normalized to the current injected within the DT guard ring.

$$1/F = \frac{I(\text{collected})}{I(\text{injected})}$$

The current that escapes outside the DT guard ring is a function of the current collected at the outer structure and the current that recombines outside the DT guard ring. Hence, if the probability that an electron recombines outside of the guard ring structure, then measured current is the probability of escape outside of the DT guard ring.

Experimental results show that the probability of escape decreases with trench depth. Assuming a functional of the power series form, the probability of escape can be expressed as

$$\frac{1}{F} = \frac{A}{(L_{\text{DT}})^B}$$

Analytical results by Watson and Voldman show that in a p-substrate $A = 0.3575$ and $B = 0.9568$. This indicates that the guard ring metric is proportional to the inverse of the depth. Inverting the expression, it can be shown that the linear relationship for F is

$$F = C(L_{\text{DT}}) + D$$

where $C = 2.7$ and $D = -0.19$ provides the best fit for the experimental results. These results show that the guard ring efficiency metric increases at approximately three-fold for every micrometer of deep trench depth in a p-substrate (Figure 6.18).

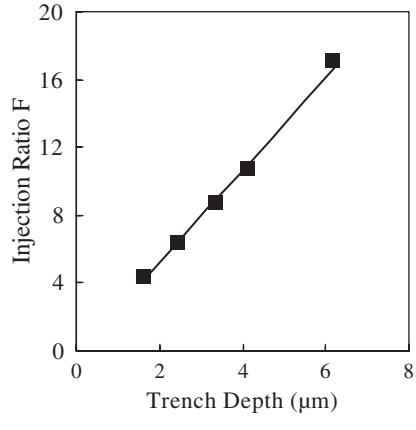

Figure 6.18 Deep trench structure guard ring injection efficiency as a function of trench depth

6.3.2 Deep Trench and Latchup

Deep trench technology can be used to reduce latchup sensitivity by placement of the deep trench structure within the lateral parasitic *pnpn* structure [31]. In semiconductors, lateral *pnpn* structures are formed by the p^+ and n^+ diffusions within the well and substrate, respectively. The latchup sensitivity of the structure is dependent on the lateral parasitic bipolar current gain of the lateral *pnp*, and lateral *npn*, as well as the well and substrate resistances. Placement of the deep trench structure on the perimeter of the *n*-well region prevents lateral current flow from the p^+ diffusion to the substrate, minimizing the lateral *pnp* bipolar current gain. Additionally, the placement of the trench structure along the *n*-well perimeter also lowers the lateral *npn* bipolar current gain. The placement of the deep trench structure leads to an increase of the *npn* avalanche multiplication. The addition of the trench structure leads to an increase in the latchup trigger voltage, as well as the latchup holding voltage and current. The latchup metric results are a function of the *n*-well doping concentration and well contact spacing, the substrate doping concentration and substrate contact spacing, and the deep trench depth (Figure 6.19).

6.3.3 Deep Trench and ESD Structures

Deep trench isolation can be used for ESD protection structures and networks. Deep trench structures can be used in a p^+/n-well diode ESD structure where the DT isolation is used to isolate the cathode structure. Deep trench isolation is placed on the sidewall of the *n*-well region. Shallow trench isolation is used to define the p^+ diffusion anode and the n^+ cathode contact regions. Shallow trench isolation is formed over the deep trench structure. The ESD p^+/n-well element is isolated from adjacent structures for density and minimizing lateral interaction. In this structure, the *n*-well-to-substrate capacitance is reduced. This element can be used as a forward-biased diode structure with the p^+ diffusion as an anode, and the *n*-well as a cathode for positive pulse events. In this case, the p^+/n-well element is influenced by the DT structure as a result of the increase in thermal resistance. With the trench isolation structure, the thermal impedance to the bulk substrate increases, leading to a higher temperature near the device surface for the same input power. From the thermal resistance models, the peak temperature will increase. An advantage of this structure is injection to the substrate is reduced and interaction laterally with adjacent elements for improved isolation, increased latchup immunity, and reduction of minority-carrier holes into the semiconductor substrate.

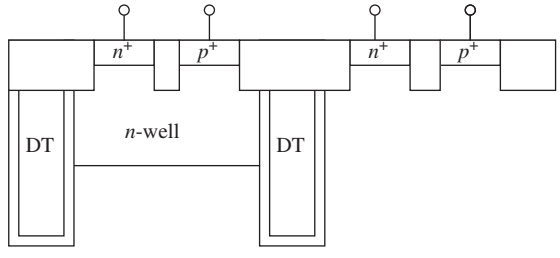

Figure 6.19 Deep trench latchup structure

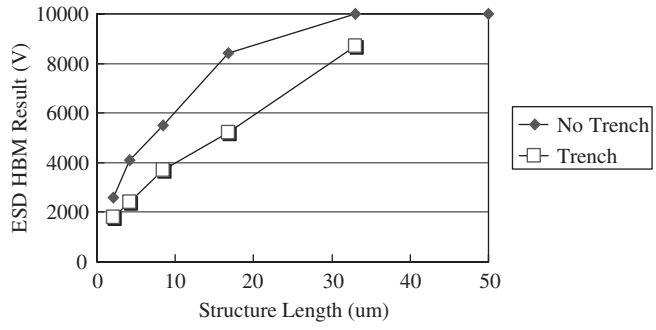

Figure 6.20 ESD sub-collector-to-substrate diode HBM results with and without deep trench

Additionally, the element can be used as a *n*-well-to-substrate diode element for negative pulse events, where the substrate serves as the anode, and the *n*-well serves as the cathode. With the trench isolation structure, the thermal impedance to the bulk substrate increases, leading to a higher temperature near the device surface for the same input power. From the thermal resistance models, the peak temperature will increase. An advantage of this structure is the reduction of interaction laterally with adjacent elements for improved isolation, density and increased latchup immunity. Experimental results (Figure 6.20) shows that ESD failure with the deep trench structure is two-fold lower than the structures without the DT boundary (in the case when the DT structure is approximately twice the depth of the collector–substrate metallurgical junction).

DT isolation can be used to form diode structures used as an isolation structure for the anode region instead of the cathode region. In this fashion, the region below the STI-defined p^+ implant and the n^+ implant is in a low-doped *p*-region. Using a sub-collector, and DT region, the sub-collector and the DT isolates the low doped *p*-region from the *p*-substrate region. In this device, the DT serves as a means to isolate the anode and cathode from the semiconductor chip substrate [34].

In Chapter 7, MOSFET drain structures and salicides will be discussed. MOSFET drain structures and salicide have significant influence on ESD protection. Since the literature has a tremendous focus on MOSFET drain structures, this will not be emphasized in depth. MOSFET drain engineering evolution as well as the transition from titanium to cobalt salicide will be discussed. A discussion on titanium properties, the transition to low resistance and the use of molybdenum in assisting the transition is briefly addressed. A focus will be maintained on thermal models on salicides relating to the self-heating effects from ESD events.

PROBLEMS

6.1. Assuming an ESD event which leads to 10 V across a junction, calculate the electric field difference between a LOCOS-bound junction and an STI-bound junction based on the S. K. Ghandi curvature relationships.

6.2. Derive the transmission probability relationship for the deep trench structure.

6.3. Derive a bipolar transistor model where there is a first *n*-well region, a trench of depth *L*, and a second well region. How does the gain relate to the trench depth and the probability of escape?

6.4. What are the reflection, absorption and transmission coefficients, assuming three types of region between a first and second surface: a *p*-contact region, a well region, or a trench region.

REFERENCES

1. S. M. Sze. *Physics of Semiconductors. Section 2.5 Junction Breakdown*. Wiley, 1981.
2. S. M. Sze and G. Gibbons. Effect of curvature on breakdown voltages in semiconductors. *Solid State Electronics*, 1966; **9**: 831.
3. S. K. Ghandi. *Semiconductor Power Devices*. New York: Wiley, 1977.
4. M. Jaffe and P. Cottrell. Electrostatic discharge protection in a 4-Mb DRAM. *Proceedings of the EOS/ESD Symposium*, 1990; 51–58.
5. A. Amerasekera and C. Duvvury. *ESD in Silicon Integrated Circuits* U. Chichesten: Wiley, 1995.
6. C. Duvvury, R. N . Rountree and, L. S. White. A summary of most effective electrostatic protection circuits for MOS memories and their observed failure modes. *Proceedings of the EOS/ESD Symposium*, 1983; 181–184.
7. Duvvury, R. Rountree, D. Baglee, A. Hyslop and L. White. ESD design considerations for ULSI. *Proceedings of the EOS/ESD Symposium*, 1985; 45–48.
8. Duvvury, R. McPhee, D. Baglee and R. Rountree. ESD protection reliability in 1 μm CMOS circuit performance. *Proceedings of the International Reliability Physics Symposium*, 1986; 199–208.
9. R. McPhee, C. Duvvury, R. Rountree and H. Domingos. Thick oxide performance under process variations. *Proceedings of the EOS/ESD Symposium*, 1986; 173–179.
10. R. Rountree and C. L. Hutchins. NMOS protection circuity. *IEEE Transactions on Electron Devices*, 1985; **ED-32**: 910–917.
11. D. Wilson, H. Domingos and M. M. Hassan. Electrical overstress in nMOS silicided devices. *Proceedings of the EOS/ESD Symposium*, 1987; 265–273.
12. H, Weston, V. Lee and T. Stanik. A newly observed high frequency effect on ESD protection utilized in a Gigahertz NMOS technology. *Proceedings of the EOS/ESD Symposium* 1992; 95–98.
13. B. Davari *et al.* A variable-size shallow trench isolation (STI) technology with diffused sidewall doping for sub-micron CMOS. *International Electron Device Meeting (IEDM) Technical Digest*, 1988; 92–95.
14. E. Adler, J. DeBrosse, S. Geissler, S. Holmes, M. Jaffe, J. Johnson, C. Koburger, J. Lasky, B. Lloyd, G. Miles, J. Nakos, W. Noble and S. Voldman. The evolution of IBM CMOS DRAM technology. *IBM Journal of Research and Development* 1995; **39** (1/2): 167–188.
15. S. Voldman, V. Gross, M. Hargrove, J. Never, J. Slinkman, M. O'Boyle, T. Scott and J. Delecki. Shallow trench isolation double-diode electrostatic discharge circuit and interaction with DRAM circuitry. *Proceedings of the EOS/ESD Symposium* 1992; 277–288.
16. S. Voldman. Shallow trench isolation double-diode electrostatic discharge circuit and interaction with DRAM circuitry. *Journal of Electrostatics*, 1993; **31**: 237–262.
17. S. Voldman and V. Gross. Scaling, optimization and design considerations of electrostatic discharge protection networks in CMOS technology. *Proceedings of the EOS/ESD Symposium*, 1993; 251–260.
18. S. Voldman, S. Furkay and J. Slinkman. Three-dimensional electrothermal simulation of electrostatic discharge protection circuits. *Proceedings of the EOS/ESD Symposium* 1994; 246–256.

19. S. Voldman *et al.* Retrograde well and epitaxial thickness optimization for shallow and deep collar MINT substrate plate trench (SPT) cell and CMOS logic technology. *Proceedings of the International Electron Device Meeting (IEDM)*, 1992; 811–814.
20. S. Voldman and G. Gerosa. Mixed voltage interface ESD protection circuits for advanced microprocessors in shallow trench and LOCOS isolation CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1994; 277–281.
21. S. Voldman, V. Gross, S. Furkay, J. Slinkman and G. Gerosa. Analysis of snubber-clamped diode string mixed voltage interface ESD protection networks for advanced microprocessors. *Proceedings of the EOS/ESD Symposium*, 1995; 43–61.
22. J. Never and S. Voldman. Failure analysis of shallow trench isolation ESD structures. *Proceedings of the EOS/ESD Symposium* 1995; 273–288.
23. S. Voldman, J. Never, S. Holmes and J. Adkisson. Line-width control effects on MOSFET ESD robustness. *Proceedings of the EOS/ESD Symposium*, 1996; 106–116.
24. S. Voldman, S. Geissler, J. Nakos and J. Pekarik. Semiconductor process and structural optimization of shallow trench isolation-defined and polysilicon-bound source/drain diodes for ESD networks. *Proceedings of the EOS/ESD Symposium*, 1997; 151–160.
25. S. Geissler, B. Porth, J. Lasky, J. Johnson and S. Voldman. A new three dimensional MOSFET gate induced drain leakage effect. *International Electron Device Meeting (IEDM) Technical Digest*, December 1991; 839–842.
26. R. Singh, D. Harame and M. Oprysko. *Silicon Germanium: Technology, Modeling and Design*, Wiley, 2004.
27. D. Walkey *et al.* Prediction of thermal resistance in trench isolated bipolar device structures. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, 1998; 207–210.
28. D. Walkey, T. Smy, C. Reimer, M. Schroter, H. Tran and D. Marchesan. Modeling thermal resistance in trench-isolated bipolar technologies including trench heat flow. *Solid State Electronics* 2002; **46**: 7–17.
29. J. S. Rieh *et al.* Structural dependence of the thermal resistance of trench-isolated bipolar transistor model. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, September 2002; 100–103.
30. S. Voldman *et al.* The influence of process and design on the ESD robustness of ESD structures and silicon germanium heterojunction bipolar transistors in a BiCMOS SiGe technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, April 2003; 347–356.
31. A. Watson, S. Voldman and T. Larsen. Deep trench guard ring structures and evaluation of the probability of minority carrier escape for ESD and latchup in advanced BiCMOS SiGe technology. *Proceedings of the Taiwan Electrostatic Discharge Conference*, 12–13 November 2003; 97–103.
32. S. Voldman. The effect of deep trench isolation, trench isolation, and sub-collector on the electrostatic discharge (ESD) robustness of radio frequency (RF) ESD STI-bound p^+/n-well diodes in a BiCMOS silicon germanium technology. *Proceedings of the 25th EOS/ESD Symposium*, 2003; 214–223.
33. S. Voldman and A. Watson. The influence of deep trench and substrate resistance on the latchup robustness in a BICMOS silicon germanium technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2004; 135–142.
34. S. S. Chen, T. Y. Chen, T. H. Tang, T. L. Hsu, H. C. Tseng, J. K. Chen and C. H. Chou. Investigation of ESD devices in 0.18 μm SiGe BiCMOS process. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 357–361.

7 Drain Engineering, Salicides and ESD

Drain engineering and salicides have a considerable influence on the ESD robustness of MOSFET transistors. In this chapter, we will focus on recent advancements in drain engineering and salicide films. Technology evolved from abrupt junctions, to low-doped drains, to extension drain implants which influenced the ESD robustness of MOSFET and gated diode structures. As the literature is extensive on the ESD results with different drain structures, the chapter will focus more on the evolutionary trend of the silicides. The evolution from titanium to cobalt silicide will also be discussed in its relevance to ESD devices and ESD engineering.

7.1 JUNCTIONS

Drain structures are the region of the transistor where significant self-heating occurs during ESD events [1–10]. The ESD current flows from the MOSFET drain-to-source region in the channel region. During the ESD event current is flowing through the drain and the source region. In this physical region, Joule heating occurs in the drain structure. The current density in the junction region can be assumed simplistically to be equal to the total current divided by the cross-sectional area of the MOSFET drain junction,

$$J \simeq \frac{I}{W_{\text{eff}}(x_{\text{J}})}$$

where W_{eff} is the effective width of the MOSFET, and x_{J} is the depth of the MOSFET source/drain junction. The Joule heating can be assumed to be equal to

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

current density times the electric field in the drain structure. This can be expressed as the square of the current density divided by the conductivity

$$\frac{1}{\sigma}\left[\frac{I}{W_{\text{eff}}X_{\text{J}}}\right]^2$$

MOSFET source/drain junctions have been optimized for the scaling of the MOSFET devices according to MOSFET constant electric field scaling theory. In MOSFET constant electric field scaling theory, the dimensions and doping concentration are scaled to provide a constant electric field across the MOSFET gate dielectric. From MOSFET constant electric field scaling, with the scaling parameter, α

$$L' = \frac{L}{\alpha}$$

$$W' = \frac{W}{\alpha}$$

$$t'_{\text{ox}} = \frac{t_{\text{ox}}}{\alpha}$$

and to maintain dimensional similitude, the junction depth must be scaled as

$$X_{\text{J}}' = \frac{X_{\text{J}}}{\alpha}$$

Assuming the ESD current is fixed, and for a fixed conductivity of the drain region, the scaling of the Joule heating in the junction will follow as

$$\frac{\dfrac{1}{\sigma}\left[\dfrac{I}{W'_{\text{eff}}X'_{\text{J}}}\right]^2}{\dfrac{1}{\sigma}\left[\dfrac{I}{W_{\text{eff}}X_{\text{J}}}\right]^2} \simeq \alpha^4$$

From this relationship, the scaling of the metallurgical junction according to MOSFET constant electric field scaling theory would lead to an increasing Joule heating in the MOSFET drain structure with a fixed ESD current and fixed drain conductivity. Hence, MOSFET drain engineering has a significant effect on the ESD robustness of a MOSFET transistor structure. Hence, the problem is as the MOSFET junction is scaled, the drain conductivity, and the film thickness of the physical film will play a role in the ESD sensitivity of transistors. In order to maintain a constant ESD robustness of MOSFET structures, a scaling relationship is needed which prevents the impact of MOSFET constant electric field scaling.

Additional to the scaling of the MOSFET junction depth, the introduction of silicide films in the MOSFET transistor to reduce MOSFET series resistance is critical. As the MOSFET junction depth is scaled, the series resistance of the MOSFET increases. Hence with the formation of the silicide film, the resistance is lowered. The introduction of the silicide has two effects for ESD. First, the current density is further increased

since the current flows through the low-resistance thin silicide film leading to higher current densities (even above the aforementioned Joule heating scaling). Additionally, lateral resistor ballasting reduces the effective width of the MOSFET structure.

7.1.1 Abrupt Junctions

MOSFET drain structures use a single implant drain structure for the source and drain implant. In an abrupt junction (Figure 7.1) the ESD current flows from the contact region along the complete junction depth of the MOSFET drain structure. Joule heating occurs in the MOSFET drain structure due to the high electric field and current from an ESD event.

The current density in the junction region can be assumed simplistically to be equal to the total current divided by the cross sectional area of the MOSFET drain junction,

$$J \simeq \frac{I}{W_{\text{eff}}(x_{\text{J}})}$$

where W_{eff} is the effective width of the MOSFET, and x_{J} is the depth of the MOSFET abrupt source/drain junction. The scaling of the MOSFET junction would follow

$$X_{\text{J}}{}' = \frac{X_{\text{J}}}{\alpha}$$

As the MOSFET abrupt junction is scaled the electric field in the MOSFET overlap region increased, leading to an increase in the hot electron sensitivity. As a result, the MOSFET abrupt junction is not able to be extended to advanced MOSFET development and has been replaced by the low-doped drain (LDD) MOSFET structure.

7.1.2 Low Doped Drains

To allow scaling of the MOSFET structure junction depth and to maintain hot electron and low-overlap capacitance objectives, the MOSFET device drain structure needs to

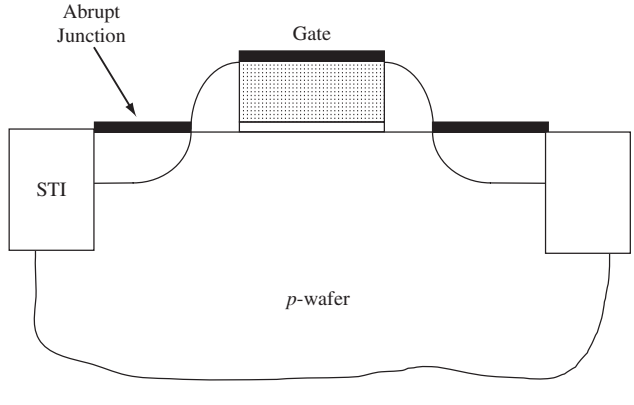

Figure 7.1 MOSFET with abrupt junction

provide MOSFET doping grading [1–4]. Abrupt junction MOSFETs require a minimum overlap region, which leads to considerable MOSFET gate-to-drain capacitance. The high electric fields also increase the hot electron sensitivity of the MOSFET transistor. The MOSFET drain structure grading was achieved using a low-doped drain (LDD) MOSFET drain structures. MOSFET LDD transistors consist of single LDD and double LDD implants. In the MOSFET LDD structures (Figure 7.2), two different dopants types are used (e.g. phosphorus and arsenic) to provide a further grading of the MOSFET drain structure. In an abrupt junction the ESD current flows from the contact region along the complete junction depth of the MOSFET drain structure.

In an LDD MOSFET structure, the current flows through the standard drain structure, followed by flow along the LDD implant region. Since the junction depth of the LDD region is less than the drain contact region, the current density in the LDD finger region is significantly higher.

Additionally, as a result of the lower doping concentration, the resistance of the LDD region is higher than the standard region. The combination of the shallower MOSFET LDD metallurgical depth and the lower doping concentration leads to an increase in the Joule heating during an ESD event. During the ESD event current is flowing through the drain and the source region where Joule heating occurs in the MOSFET drain structure. The current density in the junction region can be assumed

$$J \simeq \frac{I}{W_{\text{eff}}(X_{\text{J}})_{\text{LDD}}}$$

where W_{eff} is the effective width of the MOSFET, and x_{J} is the depth of the MOSFET LDD metallurgical junction. In the optimization of the LDD MOSFET region, the resistance is reduced by using two implants of different diffusion characteristics. With the reduction of the MOSFET LDD resistance, the Joule heating in the MOSFET finger region can be reduced. With the introduction of the MOSFET LDD structure, the ESD robustness of the MOSFET device decreases. The ESD robustness reduction with the introduction of the MOSFET LDD region in essence is a MOSFET scaling impact. To maintain constant ESD robustness, a new ESD scaling parameter is needed which maintains the constant ESD robustness of the MOSFET structure. This is

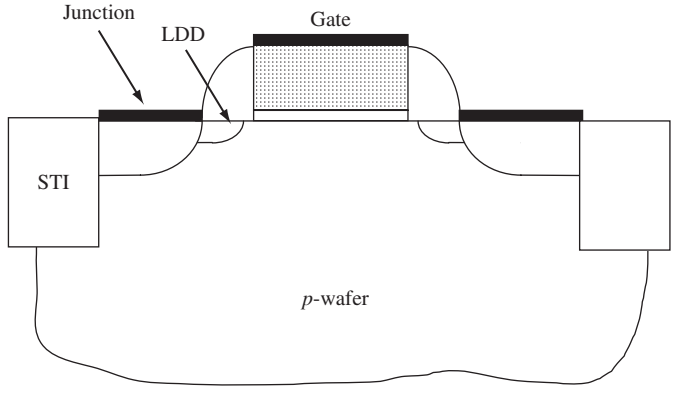

Figure 7.2 MOSFET with LDD junction

achievable by maintaining the MOSFET junction depth to allow the current to flow through the complete MOSFET source/drain region. In actuality, this solution is utilized by placement of an additional implant over the MOSFET LDD region. Comparing the LDD case with the abrupt junction, we can define the parameter as

$$\frac{(P'_{\text{diss}})_{\text{LDD}}}{P_{\text{diss}}} \simeq \frac{\dfrac{1}{(\sigma')_{\text{LDD}}}\left[\dfrac{I}{W_{\text{eff}}(X_{\text{J}})_{\text{LDD}}}\right]^2}{\dfrac{1}{\sigma}\left[\dfrac{I}{W_{\text{eff}}X_{\text{J}}}\right]^2}$$

In this expression, we can prevent the ESD robustness degradation of the MOSFET LDD device, given that, as the junction is scaled, the above expression remains a constant. To maintain ESD robustness, for the same MOSFET effective width, the product of the conductivity and the square of the junction depth must remain constant.

7.1.3 Extension Implants

Extension implants were introduced in MOSFET drain structures when the series resistance of the MOSFET LDD region impacted MOSFET performance. The MOSFET extension implants (Figure 7.3) are of a higher doping concentration compared with the MOSFET LDD structure. MOSFET extension implants for technology generations smaller than 0.25 μm technology generation when the LDD and double LDD MOSFET structures were not suitable from a performance objective. From an ESD perspective, the MOSFET extension implant region reduced the I^2R Joule heating in the MOSFET device compared with the MOSFET LDD and double LDD structures.

In this structure, the extension implants allow for the MOSFET source/drain implants to be spaced farther apart and comparatively deeper than the LDD MOSFET. From an ESD perspective, for MOSFET second breakdown, the deeper source/drain implant region will allow a larger emitter area for the lateral MOSFET bipolar transistor. As a result of the lower resistance in the MOSFET extension implant, and the

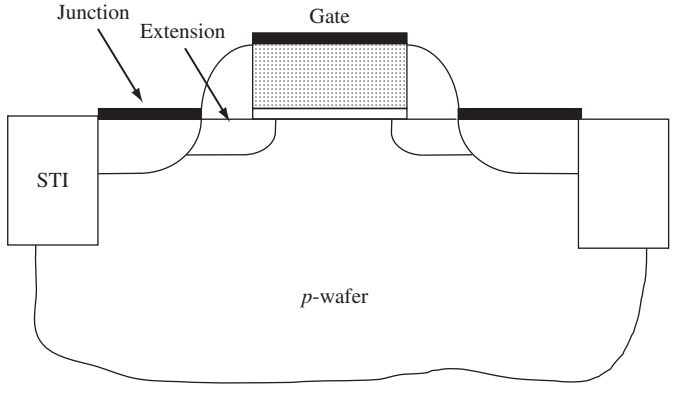

Figure 7.3 MOSFET with extension implants

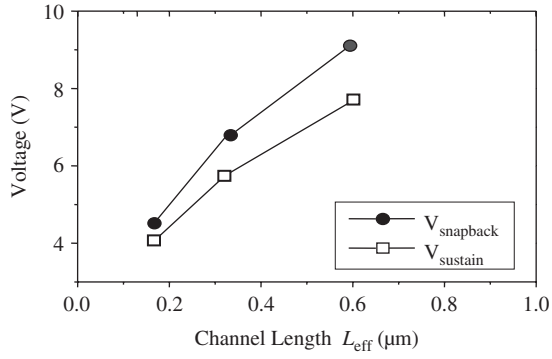

Figure 7.4 MOSFET snapback and sustaining voltage compared with technology migration

deeper MOSFET source/drain region, MOSFET second breakdown results have not significantly reduced with the introduction of the MOSFET extension implant drain structure. Figure 7.4 shows the trend of the technology with the changes in the drain structure as well as MOSFET channel length scaling.

7.2 SALICIDES AND ESD

Refractory metal films are used in semiconductors in conjunction with silicon to form silicide films [11–24]. Silicides are used in semiconductor technology to form low-resistance regions where resistance is an impediment to design or chip performance. Silicides have a considerable effect on ESD protection [1–10,25–31]. The deposition of refractory metal on the silicon surface can lead to a Schottky barrier whose current characteristics are

$$J = A^* T^2 \exp\left\{-\frac{q\phi_B}{kT}\right\}\left[\exp\left(\frac{qV}{nkT}\right) - 1\right]$$

where A^* is the Richardson constant. The barrier height of the Schottky barrier is a function of the silicide compound. It has been shown [11] that the barrier height is a function of the heat of formation, ΔH

$$\phi_B = 0.83 - 0.18(\Delta H)$$

If the doping concentration of the silicon surface is high, carriers tunnel through the barrier, leading to a low-resistance contact structure. In MOSFET structures, because of the tunneling process, silicides are used extensively. Silicides are formed on MOSFET gate structures, interconnects, contacts, and source/drain diffusions. Low-resistance gate structures allow scaling of the MOSFET channel length without the RC delay of the gate imposing a performance impact. Self-aligned silicides (salicides) are formed on MOSFET source and drain junctions to minimize the series resistance of the diffusions.

Source and drain resistance can impact the drive drain-to-source current in a MOSFET structure. In RF CMOS, lowering of the source and drain resistance is important for RF performance at high frequency.

Silicides are formed on the emitter and base structure of silicon homojunction bipolar transistors and silicon–germanium heterojunction bipolar transistors. In bipolar transistors, the emitter and collector resistance impacts the bipolar collector-to-emitter performance. The unity current gain cutoff frequency is impacted by series resistance and the delay time of an electron through the emitter, base and collector regions. For Si BJT and SiGe HBT devices, the maximum unity power gain cutoff frequency is impacted by the base resistance. Silicides can play a key role in reduction of the resistance in the emitter and base regions in high-performance transistors.

Whereas these low-resistance regions provide performance advantages in MOSFET and bipolar transistors, the proper placement and formation of silicides can influence ESD robustness. High currents in low-resistance silicide films can lead to nonuniform current distribution, current 'robbing' and impacting resistor ballasting in the structure or plurality of structures. At the same time, with proper placement or formation of silicides, current distribution can be enhanced with improved results [26, 28–31]. Understanding of silicide films at high current is important for MOSFET devices, bipolar transistors, silicided polysilicon films, electronic fuses, and other semiconductor components.

As previously discussed, in the formation of the shallow trench isolation, the objective is to achieve a planar interface between the insulator regions and the silicon areas. Etch and masking processes, and salicide formation volumetric changes, however, cause a nonplanar intersection of the isolation and silicon areas. In formation of shallow trench isolation regions, they are exposed during the etch process, leading to nonplanar STI edges where the silicon region extends above the isolation edges. In the shallow trench isolation – silicon diffusion regions, the conformally deposited refractory metal to form a silicide is deposited over the oxide regions and the STI–diffusion interface. In the formation of the silicide, the silicon laterally diffuses to react with the refractory metal, leading to interface recession. These effects can lead to perimeter diode leakage.

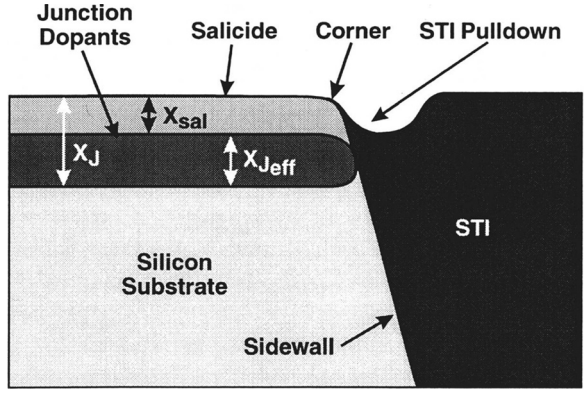

Figure 7.5 STI and salicide structure highlighting STI pull-down

The depth of the salicide can be defined as a depth X_{SAL}, and the depth of the metallurgical junction as X_J, we can define an effective junction depth X_{JEFF} [31] where

$$X_{JEFF} = X_J - X_{SAL}$$

The leakage effect and nonideal electrical characteristics can also be minimized by placement of the metallurgical junction at all significant distance from the silicide formation depth (e.g. X_{JEFF} is large enough to avoid salicide penetration into the metallurgical junction, Figure 7.5).

7.2.1 Salicide Resistance Model

Assuming a resistance model which has a first-order variation in the variable temperature of the form

$$R = R_0(1 + \alpha T)$$

let us define a relationship where the temperature in the film is a function of the thermal impedance, and the input power

$$T = \Theta_{TH} P$$

In this form, we can substitute into the resistance expression,

$$R = R_0(1 + \alpha \Theta_{TH} P)$$

Substituting in the current–resistance relationship for power [25]

$$R = R_0\left(1 + \alpha \Theta_{TH} I^2 R\right)$$

and solving for the resistance R

$$R = \frac{R_0}{1 - \alpha \Theta_{TH} I^2}$$

To evaluate the change in the resistance as a function of the current magnitude, it can be shown that

$$\frac{1}{R}\frac{dR}{dI} = \frac{2R_0 \alpha \Theta_{TH} I}{1 - \alpha \Theta_{TH} I^2}$$

This analysis assumes that the current is constant, and that the material does not change state during the self-heating process. In silicide films, the state of the material can change resistance and transformation can occur. Hence, we can assume a form where the resistance changes state during the heating process. We can assume a heuristic form

$$R = R_0(S) + R(I, V, S)$$

where $R(I,V)$ is a resistance value which is a function of the current, voltage, temperature, self-heating and material state of the film structure. Hence the validity of the equation is associated with the specific states to that power or current level.

For pulsed events, we can relate the temperature change to the input power into the device. Assuming a pulsed event, Tasca assumed the absorbed energy into the device can be of the form

$$\langle P \rangle = \langle I^2 R \rangle = I^2 \langle R \rangle = I^2 \left(\frac{R_0 + R}{2} \right)$$

where the average resistance is the average of the initial resistance and the final resistance. Equating this to an energy expression, the amount of absorbed energy is equal to the absorbed input power times the pulse width. Additionally, we can relate the temperature as the heat capacity times the change in the temperature.

$$E = C_{\text{th}} T = P\tau = I^2 \left(\frac{R_0 + R}{2} \right) \tau$$

Solving for temperature

$$T = \frac{I^2}{C_{\text{th}}} \left(\frac{R_0 + R}{2} \right) \tau$$

From this form we can substitute this into the expression for resistance [25]

$$R = R_0 \left(1 + \alpha\tau \frac{I^2}{C_{\text{th}}} \left[\frac{R_0 + R}{2} \right] \right)$$

Solving for resistance R

$$R = R_0 \frac{\left\{ 1 + \left(\frac{\alpha\tau R_0}{2C_{\text{th}}} \right) I^2 \right\}}{\left\{ 1 - \left(\frac{\alpha\tau R_0}{2C_{\text{th}}} \right) I^2 \right\}}$$

In this form, the resistance as a function of current can be evaluated in a film for d.c. and pulsed conditions.

7.2.2 Titanium Silicide

Titanium silicide is of significant importance in the semiconductor industry because of its common usage in most semiconductor technologies. Titanium silicide has wide acceptance because of its characteristics of low resistivity, good thermal stability and

self-alignment with silicon [11–17]. It is also very important in electrostatic discharge (ESD) applications because of its use on MOSFET, diodes, resistors and bipolar transistors [26, 28–31]. Titanium silicide forms on single-crystal silicon and polysilicon. A key reason why silicides are very important for ESD applications is because of its properties of low resistivity, film thickness, and the regions where silicides are formed. First, the low resistivity leads to elimination of the natural ballasting resistance within semiconductor devices, leading to current constriction and nonuniform current distribution. The low resistivity lowers the electrical and thermal stability of devices. For example, in a bipolar transistor which uses silicide on the emitter reduces the electrical and thermal stability of a circuit in a common emitter configuration. Second, the thin film thickness leads to high current density. High current densities lead to significant Joule heating in the silicide films. As the temperature in the silicide increases, the silicides can penetrate the metallurgical junction, leading to failure, or undergo thermal instability and phase transformation. A third issue is that silicides are typically formed in the regions where device performance with low resistance is desired. These same regions, in many cases are also the regions of devices which are ESD sensitive. For example, MOSFET source drain regions use $TiSi_2$ to reduce the MOSFET source/drain region. In a bipolar transistor, silicides are used in the emitter and base region. In the bipolar transistor, the emitter–base junction region is the most ESD-sensitive region of the transistor. Additionally, the design of how and where the silicide is formed can lead to nonuniform current distributions and regions of peak self-heating. With self-heating, the nature of the silicide films is modified. Hence, the nature of the film, and its thermal stability is important to provide ESD robust devices.

When the titanium silicide is self-aligned with silicon, it is referred to as Titanium salicide. The formation of $TiSi_2$ requires two silicon atoms for each deposited titanium atom. $TiSi_2$ is a polymorphic material and may exist in two states. $TiSi_2$ may exist in a orthorhombic body-centered (C49) state with 12 atoms per unit cell, or a thermodynamically favored orthorhombic face-centered (C54) state with 24 atoms per unit cell [15–18]. The phase transformation kinetics that influence which state the $TiSi_2$ resides is a function of the surface energy, film thickness, and microstructure. With technology scaling, the junction depths are scaled, requiring a scaling of the silicide film to avoid refractory metal atoms near the metallurgical junction.

The metastable C49 phase has a higher resistance (of the order of 60–90 $\mu\Omega$ cm) than the thermodynamically favored C54 phase (12–20 $\mu\Omega$ cm). It is known that initially in processing that the C49 phase of $TiSi_2$ forms first, and that the transition to the C54 phase requires a transformation process. The transformation from the C49 to C54 phase is a function of the surface energy, film thickness, microstructure [15–18]. The surface energy can be varied by adding impurities to the $TiSi_2$ or varying the substrate wafer materials. As the film thickness decreases, the transformation from the C49 to C54 phase is more difficult, with a smaller process window. If the time of formation is long, the $TiSi_2$ film undergoes undesirable agglomeration. As the silicide is formed, it undergoes a gradual transition from an initial high resistance to a low resistance where at times only a fraction of the film undergoes the transformation

$$\zeta = \frac{\rho_0 - \rho(t)}{\rho_0 - \rho_f}$$

The process of transition is similar to nucleation and growth mechanisms of incubation, induction, rapid growth and then slow completion. For thin films this can be explained by the Johnson–Mehl–Avrami equation

$$\ln(1-\zeta) = \pi\delta \int_0^t T^2 N (t-\tau)^2 \mathrm{d}\tau$$

where N is the nucleation rate, T is the growth rate, t is the process, time τ is the induction period [15–18]. From this relationship, by knowing the process time, the percentage of the transformed film can be calculated and from this knowledge, the film $TiSi_2$ resistivity can be estimated.

In a $TiSi_2$ film, a seed of the C54 phase must be present in order to initiate the nucleation process. As the line width is decreased, the probability of a C54 phase crystal structure decreases. Experimental results showed that the C54 grain size is influenced by the diffusion width on which the salicide is formed. As semiconductor devices are scaled, the ability to achieve the low-resistance state is less probable, with a decreasing process window between the C49 phase and agglomeration. Additionally, it was found that the transition from the C49 to the C54 phase is also more difficult in n^+ diffusions relative to a p^+ diffusion. In the early 1990s this was believed to be due to arsenic atoms in the n^+ diffusion influencing the transformation. In the mid-1990s it was shown that the molybdenum atom from the implanters assisted the transition from the C49 to the C54 phase. The introduction of the Mo atoms assisted in the extension of $TiSi_2$ to smaller linewidths [19,22,23].

For ESD applications silicide block masks are used on diffusions to prevent $TiSi_2$ formation in some regions of the design. Note that the reduction of the area can influence the transformation of a region to the low-resistivity C54 state.

In the development of the film, if the heating process continues, a rapid transition can occur between the transformation and the agglomeration process. After periods of heating, the $TiSi_2$ C54 phase microstructure at the grain boundaries becomes thin and begins to separate physically. As the heating continues, the interior atoms of the $TiSi_2$ C54 separate into independent regions, leading to increased series resistance.

During an ESD event, the self-heating in the salicide film can lead to agglomeration. The C54 phase microstructure separation can lead to nonuniform current distribution and instability. As the temperature of the ESD event increases, the likelihood of agglomeration also increases. The shift in the resistance can be observed as latent failure mechanisms.

When the salicide penetration is too significant or STI pull-down occurs, the leakage current in a junction can increase. Figure 7.6 is an example of STI-bound p^+/n-well diode leakage as a function of no $TiSi_2$, and two process thicknesses. In this hardware, where STI pull-down as well as a small X_{Jeff} the ESD robustness decreases with the thickness of the salicide (Figure 7.7).

Figure 7.8 demonstrates the salicide damage and crystal transformation after the ESD event on a STI p^+/n-well diode. Two regions are evident in the damage pattern.

To eliminate the impact of STI pull-down, various experiments have been performed to increase the effective junction depth. Figure 7.9 is an example of using a B11 implant and germanium pre-amorphization anneal. Disposable spacers were also attempted to increase the distance between the pull-down region and the salicide formation.

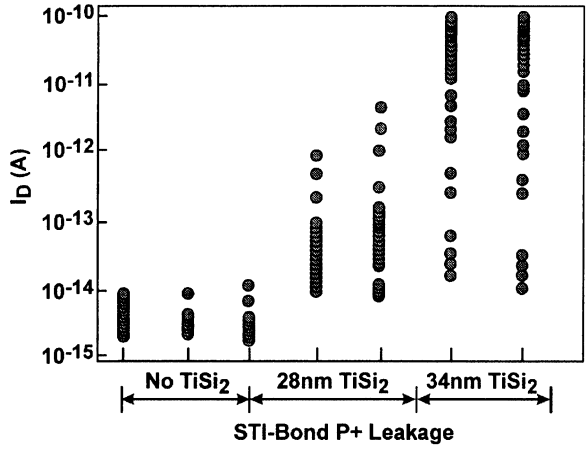

Figure 7.6 STI-bound p^+/n-well leakage with no $TiSi_2$ compared with 28 and 34 nm of $TiSi_2$

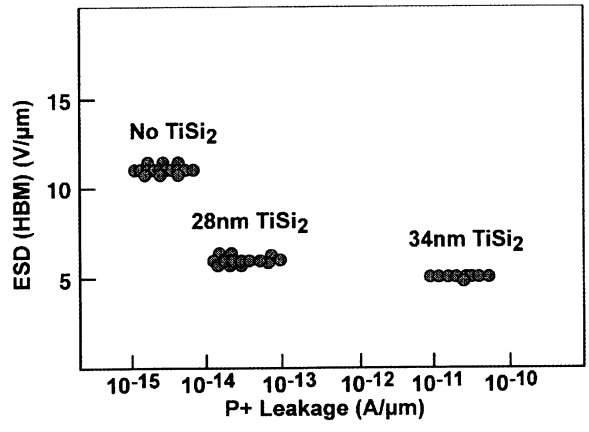

Figure 7.7 STI-bound p^+/n-well HBM ESD results with no $TiSi_2$ compared with 28 and 34 nm of $TiSi_2$

With the placement of additional implants, the ESD robustness of the diode structure can be demonstrated. Figure 7.10 shows the STI-bound p^+/n-well diode structure with B11 and an additional 20 keV implant.

With the introduction of polysilicon-bound diodes, the STI pull-down effect can be avoided. Figure 7.11 is an example of a polysilicon-bound p^+ anode structure. Figure 7.12 is an example of a second structure which has no isolation region between the p^+ and n^+ implant.

Figures 7.13 and 7.14 shows comparison of the STI-bound diode structure and the polysilicon-bound diode structures. Experimental results demonstrate the improvement of the ESD HBM robustness with the addition of the polysilicon structure.

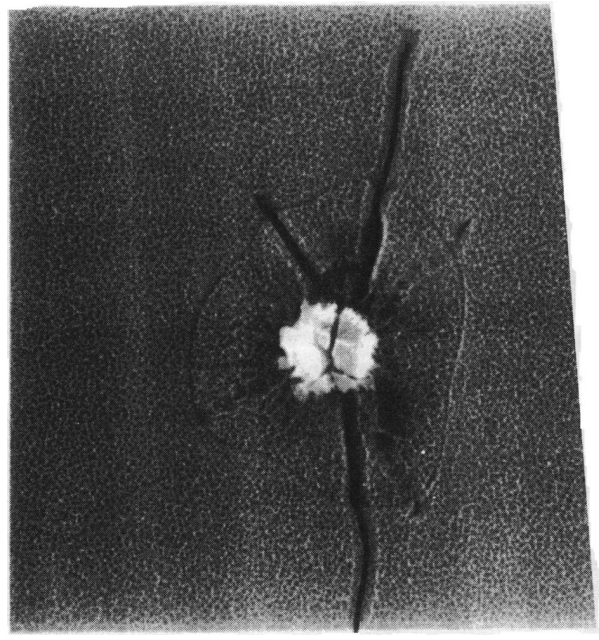

Figure 7.8 Salicide damage and crystal transformation after an ESD event on an STI p^+/n-well diode

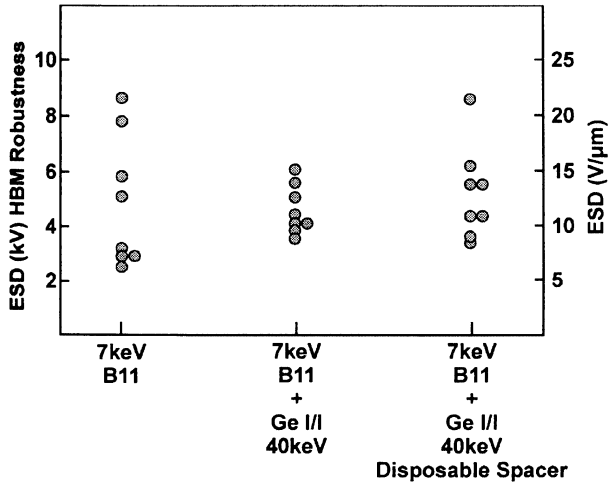

Figure 7.9 ESD HBM results of a STI-bound diode with a 40 keV Ge-PAI implant and a second Ge-PAI with a disposable spacer

ESD measurements in MOSFET structures show that the introduction of $TiSi_2$ junctions lead to a two-fold reduction in the ESD robustness of MOSFET structures [32]. This reduction has led to salicide block masks and other means of preventing ESD degradation from salicide. From this work, significant effort was devoted to eliminating silicide from LOCOS-defined MOSFET source and drain junctions.

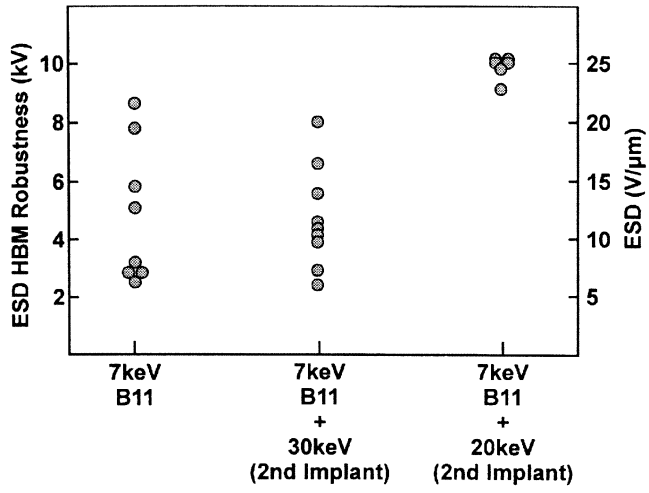

Figure 7.10 ESD HBM results for a STI-bound p^+ diode structure with a standard B11 implant and an additional deep 20 and 30 keV implant

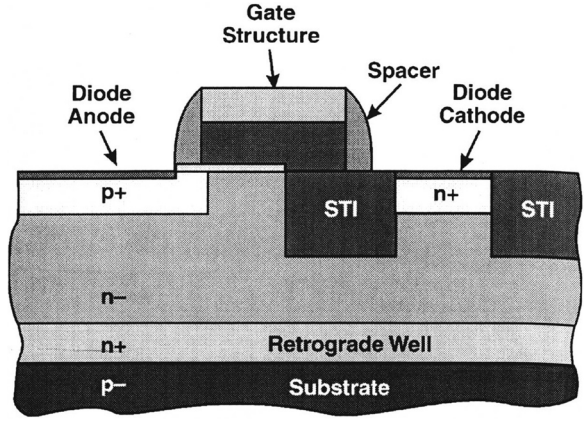

Figure 7.11 Polysilicon gate defined p^+ anode with STI-defined n^+ cathode

In shallow trench isolation (STI), diode structures demonstrated the opposite effect when no STI pull-down was evident [26, 31]. STI-bound diode structures were constructed with and without titanium disilicide. ESD measurements showed that the HBM ESD levels decreased with increased sheet resistance of the STI-bound p^+ source/drain implant. For STI-bound p^+ diodes with no salicide and p^+ sheet resistances from 20 to 28 Ω/sq., HBM ESD robustness varied from 3.1 to 3.5 kV where the ESD robustness improved for the lower sheet resistance values. For STI-bound p^+ diodes with titanium disilicide and p^+ sheet resistances from 2.8 to 3.2 Ω/sq., HBM ESD robustness varied from 3.6 to 4.2 kV where the highest ESD robustness is for the lowest sheet resistance value [26,31].

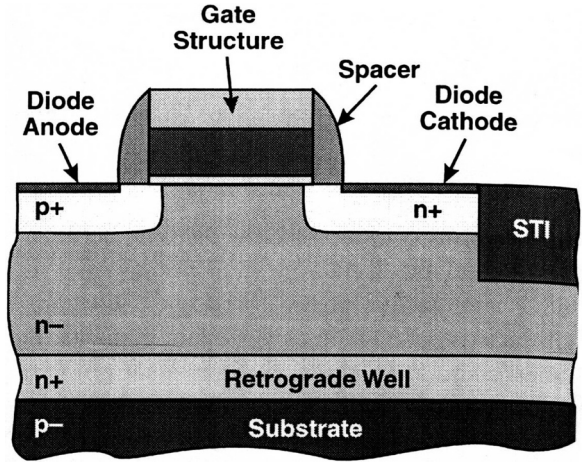

Figure 7.12 Polysilicon gate defined *p–n* diode structure

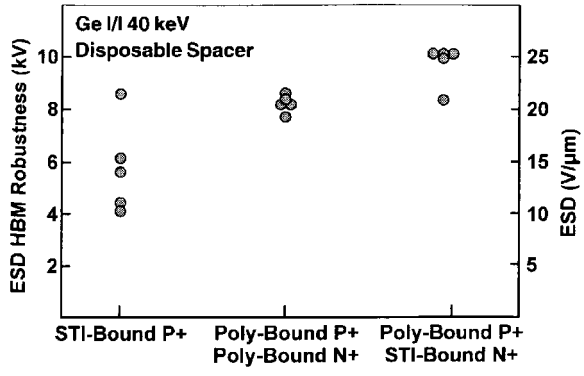

Figure 7.13 ESD HBM results for STI- and polysilicon-bound diode structures (with Ge-PAI process)

7.2.3 Molybdenum and Titanium Salicide

For titanium disilicide, the phase transformation kinetics from the C49 to the C54 state is limited by a low driving force from the small transformation enthalpy and high activation energy (e.g. greater than 5.0 eV). The metastable C49 phase has a higher resistance (of the order of 60–90 μΩ cm) than the thermodynamically favored C54 phase (12–20 μΩ cm) and the phase transition is limited by the number of C54 nuclei that exist within the silicide area. This is known as 'fineline effect' where the phase transition is limited by the number of nuclei. As the linewidth decreases, this continues to become a problem. Various approaches have been taken to initiate the full transition from the C49 to C54 state, such as rapid thermal annealing (RTA), doing with antimony (Sb), and germanium pre-amorphization implants (PAI). It was shown by R. Mann *et al.* that molybdenum (Mo) implanted prior to titanium deposition can eliminate the fineline

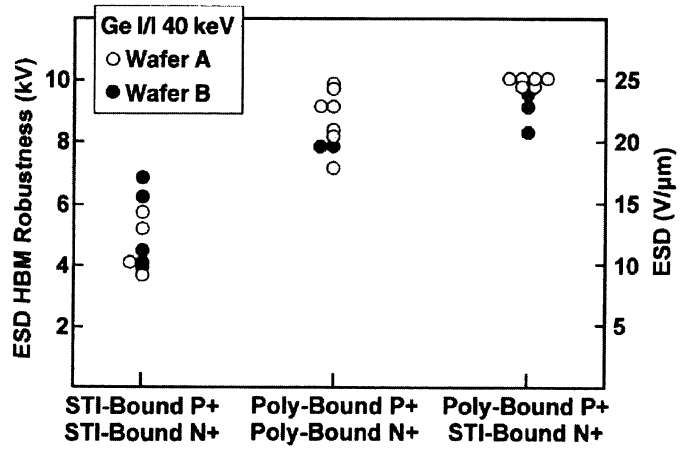

Figure 7.14 ESD HBM results for STI- and polysilicon-bound diode structures (with Ge-PAI process)

structure issue that occurs in the titanium disilicide [19,22,23]. The molybdenum implant depth is such as to be consumed in the silicon required to form the titanium disilicide. According to Mann *et al.*, the molybdenum enhances the nucleation of C54 in the C49 matrix by increasing the number of active grain triple junctions in the C49 where the C54 nucleation occurs. The molybdenum segregates into the C49 grain triple point, lowering the nucleation energy for C54 phase transformation [19,22,23].

The importance of the Mo-implanted $TiSi_2$ salicide is two-fold. First, the ability to eliminate the fineline effect from the titanium disilicide allowed extension of titanium disilicide into smaller linewidths in advanced technologies. A second issue, the Mo changes the physical stability of the titanium disilicide, leading to a different sensitivity to self-heating during and ESD event. During an ESD event, even prior to thermal runaway, the temperatures in the silicide films can exceed the activation energies of the titanium disilicide film, and can lead to shifts in the morphology and film stability. Evaluation of MOSFET transistors demonstrate the microstructure of the titanium disilicide films is changed from self-heating as after MOSFET second breakdown. Evaluation of p^+/n-well diodes also demonstrate the change in the crystalline structure beyond the diode power to failure levels.

According to Kittl *et al.*, Mo-implanted titanium disilicide films have lower thermal stability compared with standard titanium disilicide [24]. The lower thermal stability of the Mo-implanted $TiSi_2$ will lead to a higher likelihood of resistance shifts and latent damage in the ESD stressed structure. Banerjee demonstrated that the TCR of the Mo-implanted film he tested had a higher temperature coefficient of resistance (TCR) leading to a higher increase in resistance during stressing as a function of an applied TLP pulse current [25].

Voldman first demonstrated the ESD testing of Mo-implanted $TiSi_2$ films on shallow trench isolation (STI)-bound and polysilicon-bound p^+/n-well diodes [31]. The ESD results demonstrated that the transformation of the titanium disilicide more fully into the C54 state led to an improvement in both diode structures, where the polysilicon-bound diode structure showed superior ESD robustness [31].

7.2.4 Cobalt Silicides

Whereas titanium silicide is of significant importance in the semiconductor industry in 1.0–0.25 μm technology, because of the sheet resistance linewidth dependence, Cobalt silicide was needed for future technologies [13,14,24]. Titanium silicide had wide acceptance because of its characteristics of low resistivity, good thermal stability and self-alignment with silicon, but the problem of the transformation from the high resitivity orthorhombic body-centered (C49) state to the thermodynamically favored orthorhombic face-centered (C54) state at small linewidth limited its use in advanced technologies. This is critical in MOSFET gates whose physical polysilicon gate width is below 1 μm. Cobalt salicides vary from the formation of $CoSi_x$ to $CoSi_2$. Cobalt salicide formation of $CoSi_2$ undergoes transitions from [13,14].

$$Co \rightarrow Co_2Si \rightarrow CoSi \rightarrow CoSi_2$$

Using cobalt salicide, the linewidth of the polysilicon gate structures and diffusions can be reduced. It is anticipated that as the linewidth becomes small compared with the cobalt salicide grain size, nucleation of $CoSi_2$ at line edges may influence the rate of transformation. With the introduction of cobalt salicide, although it achieved objectives of low resistivity in MOSFET gate structures, it led to significant junction leakage in ultra-low junctions. Early leakage measurements by Goto *et al.* pointed out that the cobalt salicide leakage problem was driven by an area effect as opposed to perimeter effects [20,21]. Goto *et al.* claimed that cobalt salicide leakage issues occurred at the phase transition temperature from Co_2Si to CoSi as a result of CoSi spikes through the metallurgical junction area. It is well known that leakage current is a concern from the stress-induced Co_2Si/CoSi/Si triple points leading to area spiking [20,21].

With the low resistivity achieved in the cobalt salicide film, structures with salicide films will be sensitive to ESD events [25]. Cobalt salicide can be an advantage or a disadvantage for ESD elements, depending on the semiconductor device. A disadvantage of cobalt salicide is the cobalt salicide ($CoSi_2$) melting temperature of 1326 °C compared with the $TiSi_2$ melting temperature of 1500 °C. Hence, ESD elements or circuit elements whose current stressing leads to temperatures near the melting temperatures are more likely to fail at a lower critical current with cobalt salicide. Additionally, ESD structures or circuit elements whose local temperatures initiate agglomeration will also lead to salicide-induced degradation mechanism. Cobalt stability and spiking has been demonstrated to decrease with germanium pre-amorphization implants prior to the cobalt deposition.

Where cobalt salicide is an advantage was demonstrated in STI-bound p^+/n-well diode structures [31]. In a p^+/n-well diode structure, the low sheet resistance of the cobalt provides better current distribution in the diode anode. In experimental work, comparing no salicide, titanium salicide and cobalt salicide, the ESD results improved in this order, with the cobalt salicide HBM ESD results achieving double the no salicide case (Figure 7.15). Voldman showed excellent ESD robustness in cobalt salicide STI-bound p^+/n-well diode structures [31].

Cobalt salicide was formed using a 7 keV 0.18 mm deep B11 implanted STI-bound junction for a 2.5 V technology. Using germanium pre-amorphization implant

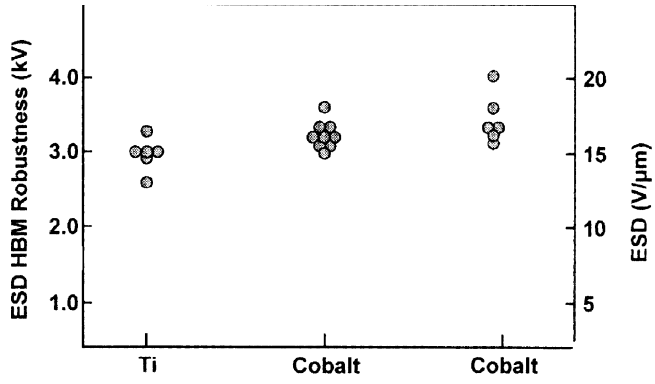

Figure 7.15 ESD HBM results of a STI-defined p^+/n-well diode with titanium and cobalt processes

(Ge-PAI), Cobalt-induced leakage increases from cobalt spiking was eliminated. In this hardware, two different cobalt processes demonstrated HBM ESD robustness superior to a $TiSi_2$ process, where the $TiSi_2$ diodes achieved approximately 3000 V HBM (see Figure 7.13). The $CoSi_x$ STI-bound diodes demonstrated results above 3000 V, with some data exceeding 4000 V [31].

In Chapter 8, dielectrics and ESD will be discussed. The chapter will primarily discuss the physics of the dielectrics. To address dielectrics in general would be quite extensive, hence a limited view of the dielectrics is shown.

PROBLEMS

7.1. Derive the temperature of the junction assuming the geometry is a parallelepiped. Assume three junctions: the abrupt junction, the LDD junction, and the extension implant. Compare the relative temperature of the three different profiles.

7.2. Derive the temperature assuming no current flows through the silicon region and only through the salicide film in a MOSFET structure. What is the power in the salicide film?

7.3. Assuming titanium and cobalt salicide films, what is the relative difference in temperature during an ESD event accounting for the thickness and sheet resistance differences?

7.4. Given the relationship for the temperature is

$$T = \Theta_{TH} P$$

substitute this expression into the Wunsch–Bell relationship where the above temperature is the failure temperature, and the power is the critical power to failure.

7.5. Given a cobalt film and a titanium salicide film have different sheet resistances, derive the ratio of the resistances using the expression

$$R = R_0 \frac{\left\{1 + \left(\frac{\alpha\tau R_0}{2C_{\text{th}}}\right) I^2\right\}}{\left\{1 - \left(\frac{\alpha\tau R_0}{2C_{\text{th}}}\right) I^2\right\}}$$

7.6. Apply the Smith–Littau relationship for a resistor element to the relationships in this chapter. How do these development differ? How are they related?

REFERENCES

1. C. Duvvury, R. N. Rountree, H. J. Steiger, T. Polgreen and D. Corum. ESD phenomena in graded junction devices. *Proceedings of the International Reliability Physics Symposium*, 1989; 71.
2. A. Amerasekera, L. V. Roozendaal, J. Bruins and F. Kuper. Characterization and modeling of second breakdown for the extraction of ESD-related process and design parameters. *IEEE Transactions on Electron Devices*, 1991; **ED-38**: 2161.
3. Y. Wei, Y. Loh, C. Wang and C. Hu. MOSFET drain engineering for ESD performance. *Proceedings of the EOS/ESD Symposium*, 1992; 143–149.
4. M. Chaine, S. Desai, C. Dunn, D. Dolby, W. Holland and T. Pekny. ESD improvements using low concentrations of arsenic implants in CMOS output buffers. *Proceedings of the EOS/ESD Symposium*, 1992; 143–149.
5. Amerasekera, M. C. Chang, J. Seitchik, A. Chatterjee, K. Mayaram and J. H. Chern. Self heating effects in basic semiconductors structures. *IEEE Transaction on Electron Devices* 1993; **ED-40**: 1836.
6. E. A. Amerasekera, W. Van den Abeelen, L. J. van Roozendahl, M. Hannemann and P. P. J. Schofield. ESD failure modes: characteristics, mechanisms, and process influences. *IEEE Transactions on Electron Devices* 1992; **ED-39**: 2.
7. A. Amerasekera and J. Seitchik. Electro-thermal behavior of deep submicron *n*MOS transistors under high current snapback ESD/EOS conditions. *International Electron Device Meeting (IEDM) Technical Digest*, 1994, 446–449.
8. R. Consiglio and T. Y. Haung. ESD reliability impact of p^+ pocket implant on double implanted *n* LDD MOSFET. *Proceedings of the EOS/ESD Symposium*, 1995; 199–207.
9. K. Bock, C. Russ, G. Badenes, G. Groeseneken and L. Deferm. Influence of well profile and gate length on the ESD performance of a fully silicided 0.25 μm CMOS technology. *Proceedings of the EOS/ESD Symposium* 1997; 308–316.
10. V. Gupta, A. Amerasekera, S. Ramaswamy and A. Tsao. ESD related process effects in mixed-voltage sub-0.5 μm technologies. *Proceedings of the EOS/ESD Symposium*, 1998; 161–169.
11. S. Murarka. *Silicides for VLSI Applications*. Orlando: Academic Press, 1983.
12. K. Maex and M. Van Rossum. *Properties of Metal Silicides*, London: INSPEC, 1995; 31–34.
13. K. Maex. Silicides for integrated circuits: $TiSi_2$ and $CoSi_2$. *Material Science and Engineering* 1993; **R11** (2–3): 53–153.
14. J. Lasky. Comparison of transformation to low resistivity phase and agglomeration of $TiSi_2$ and $CoSi_2$. *IEEE Transactions on Electron Devices*, 1991; **ED-38** (2): 262–269.

15. Clevenger, L. A., J. Harper, C. Cabral, C. Nobili, G. Ottaviani and R. Mann. Kinetic analysis of C49 $TiSi_2$ and C54 $TiSi_2$ formation at rapid annealing rates. *Journal of Applied Physics* 1992; **72**, (10): 15. 4978–4980.
16. R. Mann. The C49 to C54 $TiSi_2$ transformation in self-aligned silicide application. *Journal of Applied Physics* 1993; **73** (7): 3566–3568.
17. R. Mann, L. Clevenger and Q. Z. Hong. Phase transformation kinetics of TiSi2. *Material Research Society Symposium Proceedings*. 1993; **311**: 281–286.
18. R. Mann and L. A. Clevenger The C49 to C54 phase transformation in $TiSi_2$ thin films. *Journal of the Electrochemical Society*, 1994; **141**, (5): 1347–1350.
19. R. Mann, G. Miles, T. Knotts, D. Rakowski, L. Clevenger, J. Harper, F. D'Heurle and C. Cabral. Reduction of C54-$TiSi_2$ phase transformation temperature using refractory metal ion implantation. *Applied Physics Letters* 1995; **67** (25): 3279–3731.
20. K. Goto *et al*. A comparative study of leakage mechanism of Co and Ni salicide processes. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1998; 363–369.
21. K. Goto *et al*. Leakage mechanisms and optimized conditions of Co salicide process for deep sub-micron CMOS devices. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 1999; 449–452.
22. C. Cabral *et al*. Method of lowering the phase transformation temperature of a metal silicide. *U. S. Patent* 5,510,295, 1996.
23. R. W. Mann. Reduction of the C54 $TiSi_2$ phase transformation temperature using refractory metal ion implantation. *Applied Physics Letters* 1995; **67** (25):18. 3729–3731.
24. J. Kittl, Q. Z. Hong, C. Chao, I. C. Chen, N. Y. Yu, S. O'Brien and M. Hanratty. Salicides for 0.10 μm gate lengths: comparison of study of one step RTP Ti with Mo doping, Ti with pre-amorphization, and Co processes. *VLSI Technology Symposium Proceedings*, 1997; 103–104.
25. K. Banerjee, A. Amerasekera, J. A. Kittl and C. Hu. High current effects in silicide films for sub-0.25 μm VLSI technologies. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1998; 284–292.
26. S. Voldman. Shallow trench isolation double-diode electrostatic discharge circuit and interaction with DRAM circuitry. *Journal of Electrostatics* 1993; **31** 237–262.
27. S. Voldman *et al*. Retrograde well and epitaxial thickness optimization for shallow and deep collar MINT substrate plate trench (SPT) cell and CMOS logic technology. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 1992; 811–814.
28. J. Never and S. Voldman. Failure analysis of shallow trench isolation ESD Structures. *Proceedings of the EOS/ESD Symposium*, 1995; 273–288.
29. S. Voldman, J. Never, S. Holmes and J. Adkisson. Line-width control effects on MOSFET ESD robustness. *Proceedings of the EOS/ESD Symposium*, 1996; 106–116.
30. S. Voldman. Semiconductor diode with silicide films and trench isolation. U. S. Patent 5,629,544, May 13, 1997.
31. S. Voldman, S. Geissler, J. Nakos and J. Pekarik. Semiconductor process and structural optimization of shallow trench isolation-defined and polysilicon-bound source/drain diodes for ESD networks. *Proceedings of the EOS/ESD Symposium*, 1998; 151–160.
32. C. Duvvury. ESD reliability for advanced CMOS technologies. *Proceedings of the International Electronic Device and Material Symposium*, 1990; 265–272.

8 Dielectrics and ESD

Chapter 8 addresses ESD events and dielectrics. The ESD dielectric models of Fong and Hu on dielectric failure will be discussed. This will be followed by the work of D. Lin, addressing the distinction of the model between d.c. and pulsed phenomena. The chapter will primarily discuss the physics of the dielectrics. To address dielectrics in general would be quite exhaustive, hence a small view of the dielectrics is shown. By focusing on the analysis and physical phenomenon, the material will remain relevant for the future.

Dielectric films represent an important ESD concern in semiconductor technologies because of their ESD sensitivity, and because of scaling issues [1–10]. The effect of a high electric field in a transient environment of the gate dielectric integrity is a MOSFET scaling and ESD scaling issue. As functional performance demands increase, the oxides are scaled. At the same time, for performance objectives, ESD structures are being scaled to smaller devices, leading to an increased interest in dielectrics.

Human body model, machine model and transmission line pulse testing of receiver networks can lead to ESD failure of MOSFET dielectrics in the MOSFET receiver gate structures [11–15]. Charged device model (CDM) testing of components demonstrate that the dielectrics of receiver networks are a concern from current flowing from the substrate or well regions through the gate electrode [16–20].

From MOSFET technology scaling, the MOSFET gate thickness and channel length are scaled to maintain dimensional similarity and to maintain a constant electric field. With the MOSFET scaling, the evolution of the drain structure to maintain the same hot electron sensitivity leads to new drain and spacer structures.

Figure 8.1 is a plot of the MOSFET snapback and sustaining voltage for successive technologies.

Comparing MOSFET second breakdown and MOSFET gate dielectric breakdown as a function of technology scaling, the dielectric breakdown and the MOSFET second breakdown voltage converge as devices are scaled (Figure 8.2). As a result, both the MOSFET second breakdown and dielectric breakdown should be evaluated for prediction of ESD failure.

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

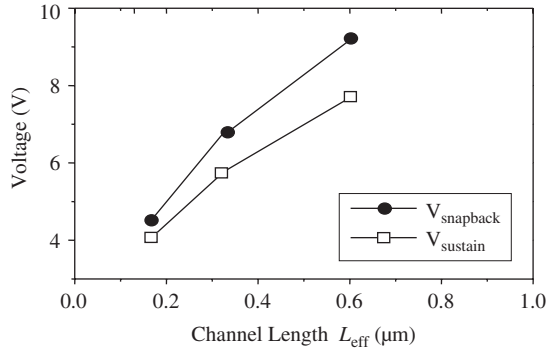

Figure 8.1 MOSFET snapback and sustaining voltage as a function of the MOSFET channel length and technology scaling

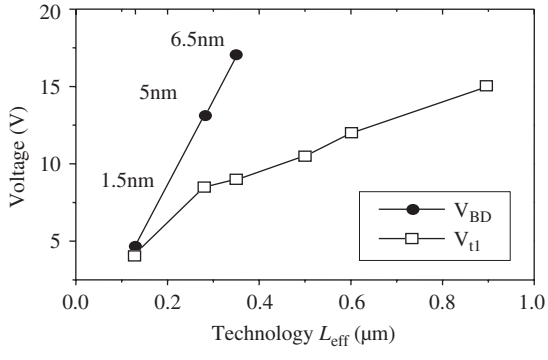

Figure 8.2 MOSFET second breakdown and MOSFET gate dielectric breakdown as a function of technology MOSFET L_{eff} and technology scaling

8.1 FONG AND HU MODEL

In this section we will look at various dielectric models, associated with ESD failure of dielectrics. For MOSFET constant electric field scaling, the dielectric film scales as

$$t' = \frac{t}{\alpha}$$

From the hole trapping model, it is believed that oxide breakdown is due to the generation of holes and the trapping of these holes at localized weak areas in the cathode region. From this assumption, Fong and Hu noted that the density of trapped holes can be expressed as

$$Q_{ot}{}^{+}(t) \propto J(E_{ox})\alpha(E_{ox})t$$

where $Q_{ot}{}^{+}$ is the density of trapped holes, $J(E_{ox})$ is the Fowler–Nordheim tunneling current density, $\alpha(E_{ox})$ is the hole generation coefficient, and t is the stress time [21]. The Fowler–Nordheim tunneling current exponential term can be expressed as

$$J \propto e^{-B/E_{ox}}$$

and the hole generation coefficient exponential term is expressed as

$$\alpha(E_{ox}) \propto e^{-H_{eff}/E_{ox}}$$

where the parameter H_{eff} is a function of the oxide thickness. Solving for time, we can express time in the form

$$t \propto \frac{Q_{ot}{}^{+}(t)}{J(E_{ox})\alpha(E_{ox})}$$

when the density of trapped holes achieves the maximum value where breakdown occurs, the time t is the breakdown time. This can then be expressed as

$$t_{BD} \propto \frac{Q_{ot}{}^{+}(t_{BD})}{J(E_{ox})\alpha(E_{ox})} \simeq Q_{ot}{}^{+}(t_{BD}) \exp\{(B + H_{eff})/E_{ox}\}$$

In this form, it was pointed out by Fong and Hu that for a fixed electric field, the different MOSFET gate dielectric thicknesses will result in different time-dependent breakdown field dependence as a result of the hole generation coefficient. In this derivation, the Fowler–Nordheim coefficient is a constant value. The expression for the breakdown can then be treated as a empirical relationship with a two-parameter form

$$t_{BD} = C_1 \exp\{C_2/E_{ox}\}$$

where it is understood that the parameter C_1 is associated with the density of trapped holes divided by the normalization coefficients of the Fowler–Nordheim equation and hole generation coefficient, and the parameter C_2 is associate with the Fowler–Nordheim coefficient B, and the hole generation term H_{eff}. In this form, plotting the log (t_{BD}) versus $1/E_{ox}$ the slope of the line is the coefficient C_2, and the intercept is related to C_1. Fong and Hu showed that the parameter C_1 (in seconds) increased from 4.4×10^{-33} to 5.8×10^{-16} and the parameter C_2 decreased from 840 to 496 for oxides from 200 to 55Å.

A failure criterion can be established using this form where failure occurs when the integral over time is such that it is greater than the parameter C_1

$$\int \mathrm{d}t \exp\{-(C_2/E_{ox}(t))\} > C_1$$

From this physical model, as the dielectric is scaled, the breakdown voltage decreases.

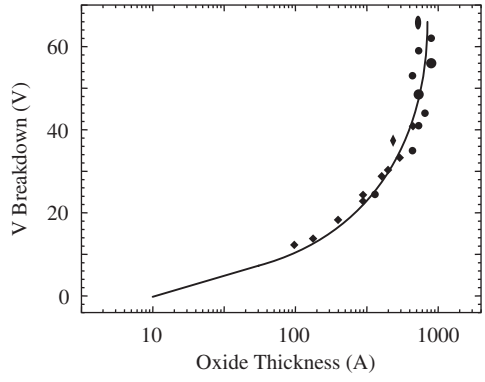

Figure 8.3 Dielectric breakdown voltage as a function of oxide thickness

Figure 8.3 shows an example of the breakdown voltage as a function of dielectric thickness for an oxide. As the oxide becomes thinner, the breakdown voltage continues to decrease.

From the Fong and Hu model, the breakdown voltage is inversely proportional to the electric field in the oxide. The form of the dielectric time to breakdown can be shown according to Figure 8.4.

The same experimental results can be represented as proportional to the oxide electric field instead of the inverse relationship. Figure 8.5 shows a sample of data from Figure 8.4 in a form proportional to the electric field across the oxide.

In ESD phenomenon, MOSFET gate receiver failures occur in the MOSFET *n*-channel transistor from positive discharge events, and in the MOSFET *p*-channel transistor from negative discharge events. MOSFET *n*-channel breakdown voltages are typically lower than the *p*-channel breakdown voltages. Figure 8.6 shows an example of the probability of gate oxide failure as a function of breakdown voltage for an *n*-channel and a *p*-channel MOSFET.

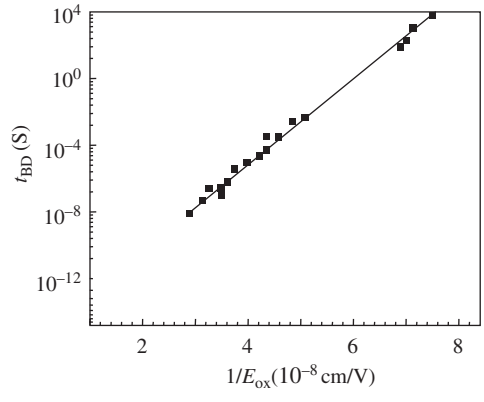

Figure 8.4 Time to breakdown t_{BD} as a function of the inverse of the oxide electric field $1/E_{ox}$

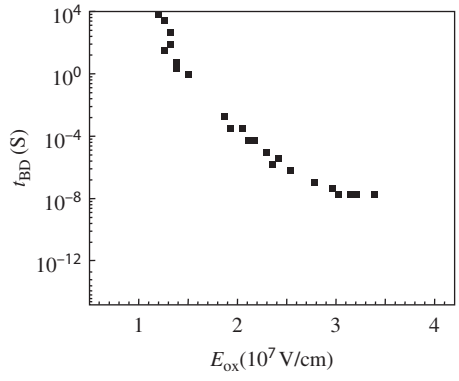

Figure 8.5 Time to breakdown t_{BD} as a function of the oxide electric field E_{ox}

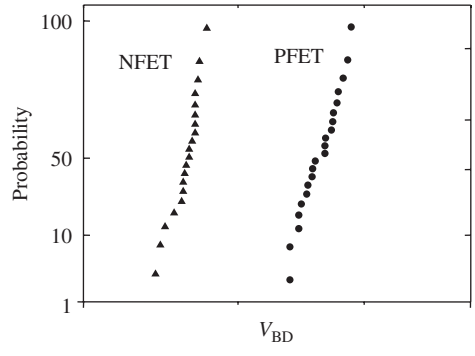

Figure 8.6 Example of probability of dielectric failure as a function of breakdown voltage for an NFET and PFET gate structure

8.2 LIN MODEL

With the scaling of thin dielectrics, the dielectric thickness and the time dynamic plays a role in the dielectric thickness. In this model, the dielectric breakdown changes with the oxide thickness and time dependence.

Assume a voltage is applied to a dielectric film. As the applied electric field is established, electrons are injected into the dielectric film. As the electrons are injected, the density of trapped holes increases with time. Assume an electron current density J, where the balance of holes is a function of the number of trapped holes from impact ionization, and the recombination cross-section of trapped holes. The hole trapping and recombination cross-section are both a function of the electric field [22]

$$e\frac{\mathrm{d}p}{\mathrm{d}t} = \alpha J - \sigma J p$$

Assuming a constant current density J, and under the assumption that the coefficients are not time dependent, then this expression can be put in the form

$$e\frac{\mathrm{d}p}{\mathrm{d}t} + (\sigma J)p = \alpha J$$

The solution for the hole density as a function of time can be obtained from a first-order linear ordinary differential equation with constant coefficients and a constant drive term. Solving the equation

$$p(t) = \frac{\alpha}{\sigma}\left[1 - \exp\left\{\frac{\sigma J}{e}t\right\}\right]$$

From the solution, it can be observed that the hole density approaches an equilibrium value of the ratio of the ionization rate and recombination cross-section, and that the time constant for this to occur is the time

$$\tau = \frac{e}{\sigma J}$$

In this model, the hole population increases until the equilibrium value where hole generation is compensated by hole recombination. In the previous dielectric model, hole recombination was not addressed. In the Fong and Hu model [21], there is a linear increase in hole generation until dielectric breakdown occurs. In this model, breakdown will occur if the saturation equilibrium level of holes is equal to or greater than the charge to breakdown density.

In the Lin model [22] it is assumed that the impact ionization rate is a function of the electric field magnitude F, but that the capture (annihilation) cross-section is assumed to be a constant. Let the impact ionization parameter be a function of the electric field, in the form of

$$\alpha = \alpha_0 e^{-H/F}$$

From Gauss's law, the derivative of the electric field is equal to the charge over the dielectric constant in the dielectric film

$$\frac{\mathrm{d}F}{\mathrm{d}x} = -\frac{e(n-p)}{\varepsilon}$$

In the approximation that the hole density is significantly higher than the electron density in the dielectric, this can be simplified to

$$\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{ep}{\varepsilon}$$

Separating the variables, we can put Gauss's law in the form

$$\int \left(\frac{\varepsilon}{ep(t)}\right) \mathrm{d}F = \int \mathrm{d}x$$

Combining the Gauss's law expression and the time-dependent hole density, the integral expression can be integrated over the dielectric region. In this expression, in order to integrate the equation, the field, and spatial terms must be separated. Putting in the field dependence of the ionization coefficient in the equation, where

$$\frac{\varepsilon\sigma}{e\alpha_0} \frac{1}{\left\{1 - \exp\left\{-\frac{\sigma J}{e} t\right\}\right\}} \int_{F_a}^{F_c} \mathrm{d}F \exp\left\{\frac{H}{F}\right\} = \int_0^{t_{ox}} \mathrm{d}x = t_{ox}$$

where the integration is over the oxide region, and the bound of the integration for the electric field is the field at the anode and cathode region. The voltage across the oxide can be expressed as the integration of the electric field across the oxide region.

$$V = \int_0^{t_{ox}} F \mathrm{d}x$$

Substituting in Gauss's law where the charge is a function of the hole population, the differential in x can be expressed as the differential in the electric field

$$V = \int_{F_a}^{F_c} F\left\{\left(\frac{\varepsilon}{e}\right) \frac{\mathrm{d}F}{p(t)}\right\} = \left(\frac{\varepsilon}{e}\right) \int_{F_a}^{F_c} \frac{F \mathrm{d}F}{p}$$

Hence, we now have two integral equations where the first is equal to the oxide field and the second is equal to the voltage across the oxide.

$$t_{ox} = \frac{\varepsilon\sigma}{e\alpha_0} \frac{1}{\left\{1 - \exp\left\{-\frac{\sigma J}{e} t\right\}\right\}} \int_{F_a}^{F_c} \mathrm{d}F \exp\left\{\frac{H}{F}\right\}$$

$$V = \left(\frac{\varepsilon}{e}\right) \int_{F_a}^{F_c} \mathrm{d}F \frac{F}{p}$$

The dimensionless variables for time, field strength, thickness and voltage are

$$\tau = \left\{\frac{\sigma J}{e}\right\} t$$

$$E = \left\{\frac{1}{H}\right\} F$$

$$T = \left\{\frac{e\alpha_0}{\varepsilon\sigma H}\right\} t_{\mathrm{ox}}$$

$$U = \left\{\frac{e\alpha_0}{\varepsilon\sigma H^2}\right\} V$$

The equation for the oxide thickness and voltage can be put in the dimensionless form

$$(1 - e^{-\tau})T = \int_{E_a}^{E_c} \mathrm{d}E \exp\{1/E\}$$

$$(1 - e^{-\tau})U = \int_{E_a}^{E_c} \mathrm{d}E E \exp\{1/E\}$$

8.3 CHARGE TO BREAKDOWN

To determine the breakdown of the dielectric film from ESD events, we must evaluate the charge to breakdown. From the normalized equations of the voltage and the oxide thickness, and the dimensionless variables for time, field strength, thickness and voltage

$$\tau = \left\{\frac{\sigma J}{e}\right\} t$$

$$E = \left\{\frac{1}{H}\right\} F$$

$$T = \left\{\frac{e\alpha_0}{\varepsilon\sigma H}\right\} t_{\mathrm{ox}}$$

$$U = \left\{\frac{e\alpha_0}{\varepsilon\sigma H^2}\right\} V$$

where the oxide thickness and voltage can be put in the dimensionless form [22]

$$(1 - e^{-\tau})T = \int_{E_a}^{E_c} \mathrm{d}E \exp\{1/E\}$$

$$(1 - e^{-\tau})U = \int_{E_a}^{E_c} \mathrm{d}EE \exp\{1/E\}$$

Initially, the hole population inside the dielectric is zero since there is no initial trapped charge and the electric field is uniform. Since the electric field at the cathode and the anode are the same, the bounds of integral are equal.

As normalized time increases, the normalized voltage U, across the electrode increases. Placing the normalized voltage in the form for small t, we can write

$$U = \frac{1}{1 - e^{-\tau}} \int_{E_a}^{E_c} \mathrm{d}EE \exp\{1/E\}$$

or in the small time limit

$$U \cong \frac{1}{\tau} \int_{E_a}^{E_c} \mathrm{d}EE \exp\{1/E\}$$

and for the normalized dielectric equation

$$T = \frac{1}{1 - e^{-\tau}} \int_{E_a}^{E_c} \mathrm{d}E \exp\{1/E\}$$

and the small time limit

$$T \cong \frac{1}{\tau} \int_{E_a}^{E_c} \mathrm{d}E \exp\{1/E\}$$

As the holes are generated, and the normalized time parameter approaches infinity, the charge in the dielectric will approach the equilibrium level if the level is below the breakdown value. The positive hole charge is equal to the hole density integrated over the dielectric region. This can be equated to the electric field relations from Gauss's law integrated over the normalized electric field.

$$Q^+(\tau) = e \int_0^{t_{ox}} p \mathrm{d}x = \varepsilon \int_{F_a}^{F_c} \mathrm{d}F$$

this can be written in the normalized variable form

$$Q^+(\tau) = e \int_0^{t_{ox}} p \mathrm{d}x = \varepsilon \int_{F_a}^{F_c} \mathrm{d}F = \varepsilon\{F_c - F_a\} = \varepsilon H\{E_c(\tau) - E_a(\tau)\}$$

Lin showed that, from first-order perturbation theory [23], the normalized integral equations can be expressed as

$$(1 - e^{-\tau})T = (E_c - E_a)e^{1/\langle E\rangle}$$

$$(1 - e^{-\tau})U = (E_c - E_a)\langle E\rangle e^{1/\langle E\rangle}$$

From the first-order perturbation form, we can relate the dimensionless average electric field to the normalized thickness, and the normalized voltage. From the two equations, by division, the average electric field is

$$\langle E\rangle = \frac{U}{T} = \frac{V}{Ht_{ox}}$$

or the voltage across the dielectric can be expressed as a function of the ionization rate constant, the average electric field and the oxide thickness.

$$V = H\langle E\rangle t_{ox}$$

From the relationship of the charge to the electric field at the cathode and the anode, we can express this as a function of the average electric field from the perturbation analysis,

$$Q^+(\tau) = \varepsilon H\{E_c(\tau) - E_a(\tau)\} = \varepsilon H\{(1 - e^{-\tau})T\exp\{-1/\langle E\rangle\}\}$$

where we can put it in un-normalized form in the oxide thickness

$$Q^+(\tau) = \frac{e\alpha_0}{\sigma}t_{ox}\{(1 - e^{-\tau})\exp\{-1/\langle E\rangle\}\}$$

In this form, the charge in the dielectric as a function of the normalized time can be expressed as a function of the ionization rate, the hole recombination cross-section, the oxide thickness, and the average electric field.

Given that the recombination time is significantly longer than the time of interest, the expression can be evaluated in the small time limit. In the small time limit, the exponential can be approximated as

$$e^{-\tau} \simeq 1 - \tau$$

$$Q^+(\tau) \cong \frac{e\alpha_0}{\sigma}t_{ox}\{(\tau)\exp\{-1/\langle E\rangle\}\}$$

where in un-normalized form, the charge in the oxide can be expressed as

$$Q^+(\tau) \cong e\alpha_0 t_{ox}\left\{\left(\frac{J}{e}t\right)\exp\{-1/\langle E\rangle\}\right\}$$

In the small τ limit, the recombination cross-section cancels out, and the charge generation in time is a linear in time. In essence, given that a time constant expansion was formed on the differential equation, where the recombination time was significantly longer than the characteristic time of the differential equation for hole generation, the equation would have been a first-order linear differential equation with only the drive term associated with the hole generation.

A general expression for the charge to breakdown is the charge needed to initiate breakdown based on the thickness, average electric field and the normalized time.

$$Q_{\mathrm{BD}}{}^{+} = \frac{e\alpha_0}{\sigma} t_{\mathrm{ox}} \{(1 - e^{-\tau}) \exp\{-1/\langle E \rangle\}\}$$

From this development, the amount of electrons injected into the dielectric can also be calculated from the hole population. This can be solved from

$$e\frac{\mathrm{d}n}{\mathrm{d}t} = J$$

$$Q(\tau)^{-} = -\int_0^t J \mathrm{d}\tau = Jt = \frac{e}{\sigma}\tau$$

Relating the first-order perturbation form for the holes

$$Q^{+}(\tau) = \frac{e\alpha_0}{\sigma} t_{\mathrm{ox}} \{(1 - e^{-\tau}) \exp\{-1/\langle E \rangle\}\}$$

where we can substitute in the electron generation for the normalized time constant, relating the hole and the electron charge population,

$$Q^{+}(\tau) = \frac{e\alpha_0}{\sigma} t_{\mathrm{ox}} \left\{ \left(1 - \exp\left\{-\frac{\sigma}{e} Q^{-}(\tau)\right\}\right) \exp\{-1/\langle E \rangle\} \right\}$$

Solving for the electron charge density,

$$Q^{-}(\tau) = -\frac{e}{\sigma} \ln\left[1 - \frac{\sigma Q^{+}(\tau)}{e\alpha_0 t_{\mathrm{ox}}} \exp\{1/\langle E \rangle\}\right]$$

Letting the breakdown criteria be associated with the measured injected current and the time to breakdown, we can obtain the charge to breakdown associated with the electron charge, as

$$Q^{-}{}_{\mathrm{BD}} = Jt_{\mathrm{BD}}$$

Substituting this into the hole charge density, we can obtain the charge to breakdown for the holes as

$$Q_{\mathrm{BD}}{}^{+} = \frac{e\alpha_0}{\sigma} t_{\mathrm{ox}} \left\{ \left(1 - \exp\left\{-\frac{\sigma}{e} Q_{\mathrm{BD}}{}^{-}\right\}\right) \exp\{-1/\langle E \rangle\} \right\}$$

and

$$Q_{\mathrm{BD}}{}^{-} = -\frac{e}{\sigma}\ln\left[1 - \frac{\sigma Q_{\mathrm{BD}}{}^{+}}{e\alpha_0 t_{\mathrm{ox}}}\exp\{1/\langle E\rangle\}\right]$$

8.4 CRITICAL DIELECTRIC THICKNESS

For ESD analysis of dielectrics, the thickness dependence of the dielectric breakdown strength is a parameter of interest. This can be obtained from the charge to breakdown formulation [21–23]. The charge to breakdown for the hole carriers and the injected electrons can be expressed as

$$Q_{\mathrm{BD}}{}^{+} = \frac{e\alpha_0}{\sigma} t_{\mathrm{ox}}\{(1 - e^{-\tau})\exp\{-1/\langle E\rangle\}\}$$

$$Q_{\mathrm{BD}}{}^{-} = -\frac{e}{\sigma}\ln\left[1 - \frac{\sigma Q_{BD}{}^{+}}{e\alpha_0 t_{\mathrm{ox}}}\exp\{1/\langle E\rangle\}\right]$$

The charge to breakdown expression can be put in the form

$$Q_{\mathrm{BD}}{}^{+} = \frac{e\alpha_0}{\sigma}(1 - e^{-\tau})[t_{\mathrm{ox}}\{\alpha_0\exp\{-1/\langle E\rangle\}\}]$$

or

$$N_{\mathrm{BD}}{}^{+} = \left[\frac{\alpha_0}{\sigma}(1 - e^{-\tau})\right][t_{\mathrm{ox}}\{\alpha_0\exp\{-1/\langle E\rangle\}\}]$$

The number of holes to breakdown can be then expressed as a first term associated with the equilibrium level and the time response to achieve the equilibrium level. Note that this first term is not a condition of breakdown, but the result of the competition between the ionization and recombination.

The second term is associated with the ionization rate and a spatial position. Lin points out that this term is associated with a condition of breakdown, analogous to breakdown derivations of the form

$$\int \alpha dx = K$$

where K is a constant. As an example, the condition for avalanche multiplication in a junction space charge region is

$$M = \frac{1}{1 - \int_{x_{\mathrm{a}}}^{x_{\mathrm{c}}} \alpha \mathrm{d}x}$$

Hence, avalanche breakdown occurs when the integral from the anode to the cathode over the integral of the ionization rate is equal to unity.

$$\int_{x_a}^{x_c} \alpha \mathrm{d}x = 1$$

By analogy, we can address the expression as

$$\{\alpha_0 \exp(-1/\langle E \rangle)\} t_{ox} = K$$

or rearranging, we can have

$$\alpha_0 \{t_{ox} \exp(-1/\langle E \rangle)\} = K$$

Hence, for a constant charge to breakdown and constant time τ, we can express this as a breakdown condition with a average electric field breakdown value, F_{BD}

$$t_{crit} = t_{ox} \exp\left\{-\frac{1}{\langle E_{BD} \rangle}\right\} = t_{ox} \exp\left\{-\frac{H}{\langle F_{BD} \rangle}\right\}$$

where

$$\alpha_0 t_{crit} = K$$

From this form, solving for the average electric field at breakdown,

$$\langle F_{BD} \rangle = \frac{H}{\ln\left\{\frac{t_{ox}}{t_{crit}}\right\}}$$

Letting

$$\langle F_{BD} \rangle = \frac{V_{BD}}{t_{ox}}$$

then

$$\langle V_{BD} \rangle = \frac{H t_{ox}}{\ln\left\{\frac{t_{ox}}{t_{crit}}\right\}} = \frac{H}{\frac{1}{t_{ox}} \ln\left\{\frac{t_{ox}}{t_{crit}}\right\}}$$

8.5 ESD PULSE EVENTS AND CHARGE TO BREAKDOWN DIELECTRIC MODEL

ESD events are a time-dependent transient phenomenon. The general first-order differential equation with variable coefficients and a time-dependent current drive term is needed for evaluation of the charge to breakdown.

Assume the ESD source is a capacitor element. This can represent the machine model, cassette model, or a limit of the human body model. Given a capacitor element C and area of the gate dielectric A, a current I from the source, and a current density J which is the current through the dielectric.

$$-C\frac{\mathrm{d}V}{\mathrm{d}t} = I = JA$$

From this form

$$J(t) = -\frac{C}{A}\frac{\mathrm{d}V(t)}{\mathrm{d}t}$$

From the general solution of the hole generation as a function of time, we can substitute in the current density from the above relationship to determine the integration factor

$$p(t) = \frac{\displaystyle\int^{t}\left\{\frac{\alpha(t')J(t')}{e}\exp\left[\int^{t''}\frac{\sigma J(t'')}{e}\mathrm{d}t''\right]\right\}\mathrm{d}t'}{\displaystyle\exp\left[\int^{t}\frac{\sigma J(t')}{e}\mathrm{d}t'\right]}$$

Solving for the integrating factor by substitution of the current density relationship to the terminal source variables C and V

$$\int^{t}\frac{\sigma J(t')}{e}\mathrm{d}t' = \frac{\sigma}{e}\int^{t}J(t')\mathrm{d}t' = \frac{\sigma}{e}\int^{t}\left\{-\frac{C}{A}\frac{\mathrm{d}V(t')}{\mathrm{d}t'}\right\}\mathrm{d}t' = -\frac{\sigma C}{eA}\int^{V}\mathrm{d}V$$

Solving for the areal trapped hole density

$$Q^{+} = e\int_{0}^{t_{\mathrm{ox}}}p(x,\infty)\mathrm{d}x$$

Using the voltage integral form of the integration factor, the area trapped hole density is equal to

$$Q^{+} = t_{\mathrm{ox}}\frac{C}{A}\int^{V_0}\alpha\exp\left\{-\frac{\sigma C}{eA}V\right\}\mathrm{d}V$$

In this form, it can be observed that the charge is proportional to the charge of the source and the oxide thickness, and the rate of the charging in the integral is a function of the ionization rate, and of an exponential factor whose time constant is the product of the number of electrons and the recombination cross-section.

Under the assumption that the exponential factor is small

$$Q^{+} \simeq t_{\mathrm{ox}}\frac{C}{A}\int^{V_0}\alpha\left[1-\frac{\sigma CV}{eA}\right]\mathrm{d}V$$

Given that $V \ll eA/\sigma C$, and substituting in the relationship for the ionization coefficient, we obtain

$$Q^+ \simeq t_{ox}\frac{C}{A}\int^{V_0} \alpha[1]\mathrm{d}V = t_{ox}\frac{C\alpha_0}{A}\int_0^{V_0} \exp\left\{-\frac{Ht_{ox}}{V}\right\}\mathrm{d}V$$

With a transformation of variables, this can be put in the form

$$Q^+ = C\{H\alpha_0\}\left\{\frac{{t_{ox}}^2}{A}\right\}\int_0^{H_{tox}/V_0} \exp\left\{-\frac{1}{x}\right\}\mathrm{d}x$$

In this form, it can be seen that the charge is a function of the ESD source capacitor, the ionization rate parameters, and the geometric variables of the charging dielectric. The integral expression is a function of the ratio of $H\,t_{ox}$ to the charging voltage source of the ESD event.

8.6 TRANSIENT PULSE EVENTS AND CHARGE TO BREAKDOWN DIELECTRIC MODEL

ESD events, such as human body model, machine model, and charged device model, are high-current pulses which are time dependent. These events vary in magnitude and can have different polarities. In the previous sections, the electron current density J was assumed constant and of single polarity. In the case of a positive voltage on the gate structure, the electrons are attracted to the positive electrode.

To evaluate the charge to breakdown in the environment of a transient pulse, we must address both a time-varying current drive term, as well as a time-dependent ionization rate. Assume a voltage is applied to a dielectric film. As the applied electric field is established, electrons are injected into the dielectric film. As the electrons are injected, the density of trapped holes increases with time. Assume an electron current density J, where the balance of holes is a function of the number of trapped holes from impact ionization, and the recombination cross-section of trapped holes. The hole trapping and recombination cross-section are both a function of the electric field.

Generalizing to include time-dependent current source and ionization rates, the ordinary differential equation becomes

$$e\frac{\mathrm{d}p(t)}{\mathrm{d}t} = \alpha(t)J(t) - \sigma J(t)p(t)$$

Assuming a time-dependent current density J, and under the assumption that the ionization rate is time dependent, but the recombination cross-section is not time dependent, then this expression can be put in the form

$$\frac{\mathrm{d}p}{\mathrm{d}t} + \left(\frac{\sigma J(t)}{e}\right)p = \frac{\alpha(t)}{e}J(t)$$

The hole density as a function of time can be solved as a the general first-order ordinary differential equation with variable coefficients and a time-dependent drive term.

Solving the equation, the integrating factor is

$$\mu(t) = \exp\left[\int^{t} \frac{\sigma J(\chi)}{e} \mathrm{d}\chi\right]$$

then the solution to the general first-order ordinary differential equation is

$$p(t) = \frac{1}{\mu(t)}\left[\int^{t} \mu(s)\left\{\frac{\alpha(s)J(s)}{e}\right\}\mathrm{d}s + C\right]$$

and we can solve for the hole density as a function of time as

$$p(t) = \frac{\displaystyle\int^{t}\left\{\frac{\alpha(t')J(t')}{e}\exp\left[\int^{t''}\frac{\sigma J(t'')}{e}\mathrm{d}t''\right]\right\}\mathrm{d}t'}{\displaystyle\exp\left[\int^{t}\frac{\sigma J(t')}{e}\mathrm{d}t'\right]}$$

From Gauss's law, the derivative of the electric field is equal to the charge over the dielectric constant in the dielectric film

$$\frac{\mathrm{d}F}{\mathrm{d}x} = -\frac{e(n-p)}{\varepsilon}$$

In the approximation, that the hole density is significantly higher than the electron density in the dielectric, this can be simplified to

$$\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{ep}{\varepsilon}$$

Separating the variables, we can put Gauss's law in the form,

$$\int\left(\frac{\varepsilon}{ep(t)}\right)\mathrm{d}F = \int \mathrm{d}x$$

where

$$p(t) = \frac{\displaystyle\int^{t}\left\{\frac{\alpha(t')J(t')}{e}\exp\left[\int^{t''}\frac{\sigma J(t'')}{e}\mathrm{d}t''\right]\right\}\mathrm{d}t'}{\displaystyle\exp\left[\int^{t}\frac{\sigma J(t')}{e}\mathrm{d}t'\right]}$$

Combining the Gauss's law expression and the time-dependent hole density, the integral expression can be integrated over the dielectric region. The voltage across the oxide can be expressed as the integration of the electric field across the oxide region.

$$V = \int_0^{t_{ox}} F \mathrm{d}x$$

Substituting in Gauss's law where the charge is a function of the hole population, the differential in x can be expressed as the differential in the electric field

$$V = \int_{F_a}^{F_c} F\left\{\left(\frac{\varepsilon}{e}\right)\frac{\mathrm{d}F}{p(t)}\right\} = \left(\frac{\varepsilon}{e}\right)\int_{F_a}^{F_c}\frac{F\mathrm{d}F}{p}$$

Hence, we now have two integral equations where the first is equal to the oxide field and the second is equal to the voltage across the oxide.

$$t_{ox} = \int \mathrm{d}x = \int_{F_a}^{F_c}\frac{\varepsilon}{ep(t)}\mathrm{d}F$$

$$V = \left(\frac{\varepsilon}{e}\right)\int_{F_a}^{F_c}\mathrm{d}F\frac{F}{p}$$

with

$$p(t) = \frac{\displaystyle\int^{t}\left\{\frac{\alpha(t')J(t')}{e}\exp\left[\int^{t''}\frac{\sigma J(t'')}{e}\mathrm{d}t''\right]\right\}\mathrm{d}t'}{\displaystyle\exp\left[\int^{t}\frac{\sigma J(t')}{e}\mathrm{d}t'\right]}$$

The dimensionless variables for time, field strength, thickness and voltage are

$$\tau = \left\{\frac{\sigma J}{e}\right\}t$$

$$E = \left\{\frac{1}{H}\right\}F$$

$$T = \left\{\frac{e\alpha_0}{\varepsilon\sigma H}\right\}t_{ox}$$

$$U = \left\{\frac{e\alpha_0}{\varepsilon\sigma H^2}\right\}V$$

8.7 ULTRA-THIN DIELECTRICS

With oxide scaling, the issue of current conduction, temperature effects, gate oxide tunneling, soft breakdown and the gate oxide reliability become an important part of

the discussion of the oxide response to the applied voltage [24–33]. As the oxides become thinner, soft breakdown becomes more likely in the 2–7 nm gate dielectric thickness regions. Weir *et al*. [27] showed that there is an increasing incidence of soft breakdown. As the oxides are scaled to 2 nm thickness, the ability to determine breakdown based on voltage drops or current spikes were not detectable, and the only 'failure' indication was based on increase in the noise.

Experimental work from Rosenbaum *et al*. for oxides in the 2.2–4.7 nm range using 1 ns VF-TLP pulse demonstrate that the oxide data continue to follow a $1/E$ model for time-dependent breakdown [29]. Experimental results show that the data for three different oxide thicknesses overlay in a linear fashion on a log (t_{BD})–$1/E_{ox}$ plot. The experimental work showed that the ESD-induced trap state generation was also found to be pulse-width-dependent [29]. It is postulated that the pulse-width-dependent trap state generation is associated with a temperature acceleration. As oxides continue to scale, the ESD concerns of these structures will continue to be an issue.

From the work of Weir *et al*. [33], breakdown is a statistical issue where, as the gate oxide area is reduced, the probability of failure decreases. Breakdown of oxides from transmission line pulse events is a statistical process. The larger the gate oxide, the higher the probability of a defect initiating the breakdown process. For example, as the area of the gate oxide decreases, the probability of failure decreases. This can be represented according to Figure 8.7.

The breakdown voltage depends on the area, where

$$V_1 = V_2 - \left(\frac{1}{V_{acc}}\right)\left(\frac{1}{\beta}\right)\log\left(\frac{A_1}{A_2}\right)$$

where V_1 is the breakdown voltage for area A_1, and V_2 is the breakdown voltage for area A_2. This can be expressed as

$$V_1 = V_2 - \left(\frac{1}{V_{acc}}\right)\left(\frac{1}{\beta}\right)\log\left(\frac{\ln(1-f_2)}{\ln(1-f_1)}\right)$$

where f_1, and f_2 are associated with the failure fraction.

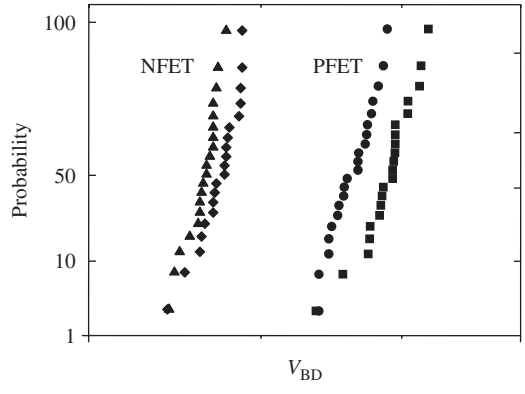

Figure 8.7 Example of probability of dielectric failure as a function of breakdown voltage for an NFET and PFET gate for two different area gates

In Chapter 9, interconnects and ESD are discussed. The wiring levels, vias and interlevel dielectric films will be discussed. The physics of the interlevel dielectrics and ESD is relevant to the current on dielectrics, but with thicker films. The primary focus on the next chapter will be the interconnect itself, and how the dielectric thermal properties influence the interconnect robustness. Interconnects and ESD will be discussed in terms of aluminum interconnects, copper interconnects, and copper/low-k interlevel dielectric systems relevant to today's semiconductor chips.

PROBLEMS

8.1. Given the dielectric thickness scales with each technology generation according to the MOSFET constant electric field scaling parameter α, derive the breakdown voltage assuming a d.c. model.
8.2. Given the dielectric thickness scales with each technology generation according to the MOSFET constant electric field scaling parameter α, derive the breakdown voltage assuming a transient pulse model.
8.3. What is the trigger voltage needed for an ESD network under the constraint that each technology generation dielectric thickness is scaled according to the MOSFET constant electric field scaling parameter?
8.4. What is the necessary charge and current allowed by an ESD network under the constraint that each technology generation dielectric thickness is scaled according to the MOSFET constant electric field scaling parameter?
8.5. Given that a transmission line pulse rise time t_{rise}, is an integer multiple of electron transport time across a dielectric film, where the transport is ballistic, relate the Lin model to both the rise time and the electron transport time. Derive the relationship assuming tunneling transport.
8.6. Derive the ratio of the Fong and Hu model to the Lin transient model for Q_{BD}.

REFERENCES

1. T. H. DiStefano and M. Shatzkes. Dielectric instability and breakdown in SiO_2 thin films. *Journal of Vacuum Science Technology* 1976; **13**: 50–54.
2. T. H. DiStefano and M. Shatzkes. Impact ionization model for dielectric instability and breakdown. *Applied Physics Letters* 1974; **25**: 685–687.
3. J. J. O'Dwyer. Theory of conduction in dielectrics. *Journal of Applied Physics* 1969; **39**: 3887–3890.
4. D. J. DiMaria, D. Arnold and E. Cartier. Impact ionization and positive charge formation in silicon dioxide films on silicon. *Applied Physics Letters* 1992; **60**: 2118–2120.
5. D. Arnold, E. Cartier and D. J. DiMaria. Acoustic phonon runaway and impact ionization by hot electrons in silicon dioxide. *Physical Review B* 1992; **45**: 1477–1480.
6. M. Fischetti, D. J. DiMaria, S. D. Brorson, T. N. Theis and J. R. Kirtley, Theory of high field electron transport in silicon dioxide. *Physical Review B*, 1985; **31**: 8124–8142.

7. K. F. Schuegraf and C. Hu. Hole injection oxide breakdown model for very low voltage lifetime extrapolation. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1993; 7–12.
8. N. Klein and P. Solomon. Current runaway in insulators affected by impact ionization and recombination. *Journal of Applied Physics* 1976; **47**: 4364–4372.
9. Y. Nissan-Cohen, J. Shappir and D. Frohman-Bentchkowsky. High yield and current induced positive charge in thermal SiO_2 layers. *Journal of Applied Physics*, 1985; **57**: 2830–2839.
10. I. C. Chen, S. Holland and C. Hu. Electron-trap generation by recombination of electrons and holes in SiO_2. *Journal of Applied Physics* 1987; **61**: 4544–4548.
11. A. Amerasekera and D. Campbell. ESD pulse and continuous voltage breakdown in MOS capacitors structures. *Proceedings of the EOS/ESD Symposium*, 1986; 208–213.
12. D. L. Lin and T. Welsher. From lightning to charged-device model electrostatic discharges. *Proceedings of the EOS/ESD Symposium*, 1992; 68–75.
13. M. Bridgewood and Y. Fu. A comparison of threshold damage processes in thick field oxide protection devices following square pulse and human body model injection. *Proceedings of the EOS/ESD Symposium*, 1988; 126–136.
14. M. Bridgewood and R. Kelly. Modeling the effects of narrow impulsive overstress on capacitive test structures. *Proceedings of the EOS/ESD Symposium*, 1985; 84–91.
15. M. J. Tuncliffe, V. M. Dwyer and D. S. Campbell. Experimental and theoretical studies of EOS/ESD oxide breakdown in unprotected MOS structures. *Proceedings of the EOS/ESD Symposium*, 1990; 162–168.
16. *CDM - ESD Sensitivity Testing: Charged Device Model (CDM) – Component Level ESD Association Standard*, D5.3, 1997.
17. Y. Kitamura, H. Kitamura, K. Nakanishi and Y. Shibuya. Breakdown of thin gate-oxide by application of nanosecond pulse as ESD test. *Proceedings of International Test and Failure Analysis (ITSFA)* 1989; 193–199.
18. T. J. Maloney. Designing MOS inputs and outputs to avoid oxide failure in the charged device model. *Proceedings of the EOS/ESD Symposium*, 1988; 220.
19. R. G. Renninger. Mechanism of charged device electrostatic discharge. *Proceedings of the EOS/ESD Symposium*, 1991; 127.
20. R. G. Renninger, M-C. Jon, D. L. Lin, T. Diep and T. L. Welsher. A field-induced charged device model simulator. *Proceedings of the EOS/ESD Symposium*, 1989; 59–71.
21. Y.-P. Fong and C. Hu. The effects of high electric field transients on thin gate oxide MOSFETs. *Proceedings of the EOS/ESD Symposium*, 1987; 252–257.
22. D. L. Lin. ESD sensitivity and VLSI technology trends: thermal breakdown and dielectric breakdown. *Proceedings of the EOS/ESD Symposium*, 1993; 73–82.
23. D. L. Lin. Electron multiplication in solids. *Physical Review B*, 1979; **20**: 5238–5245.
24. R. Degraeve *et al*. Temperature acceleration of oxide breakdown and its impact on ultra-thin gate oxide reliability. *Proceedings of the VLSI Technical Symposium*, 1999; 59–60.
25. R. Moamzzami *et al*. Temperature acceleration of time-dependent dielectric breakdown. *IEEE Transactions Electron Devices* 1989; **ED-36** (11): 2262–2265.
26. T. Nigam, S. Martin and D. Abusch-Magder. Temperature dependence and conduction mechanism after analog soft-breakdown. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 417–423.
27. B. Weir *et al*. Ultra-thin dielectrics: they breakdown but do they fail? *International Electron Device Meeting (IEDM) Technical Digest*, 1997; 73–76.
28. J. Stathis and D. J. DiMaria. Reliability projections for ultra-thin oxides at low voltages. *International Electron Device Meeting (IEDM) Technical Digest*, 1998; 167–170.
29. J. Wu, P. Juliano and E. Rosenbaum. Breakdown and latent damage of ultra-thin oxides under ESD stress conditions. *Proceedings of the EOS/ESD Symposium*, 2000; 287–295.

30. B. Linder, J. H. Stathis, D. J. Frank, S. Lombardo and A. Vayshenker. Growth and scaling of oxide conduction after breakdown. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 402–405.
31. M. A. Alam and R. K. Smith. A phenomenological theory of correlated multiple soft-breakdown events in ultra-thin oxides. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 406–412.
32. K. P. Cheung. Soft breakdown in thin gate oxide—a measurement artifact. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003; 432–437.
33. B. Weir *et al*. Gate dielectric breakdown: a focus on ESD protection. *Proceedings of the International Reliability Physics Symposium (IRPS)*, ESD Session, 2004; 399–404.

9 Interconnects and ESD

In this chapter, interconnects and their importance in ESD robustness in semiconductor components will be discussed. The chapter will discussion the evolution of interconnects from aluminum-based interconnects, to copper interconnects, and copper/low-k interlevel dielectric interconnect systems. These models will be extended to address voids, and insulator 'fill shapes.' The chapter will continue to discuss the role of microstructure changes and material changes as a result of ESD events. A brief discussion will also include the issue of insulator thermal and mechanical stress effects from ESD events. Although there is a number of interconnect models, the focus will be on a few of these. The methods developed in Chapters 2–5 are all applicable to the case of the interconnect because the interconnect and via are a parallelepiped structures. Hence the models of Wunsch–Bell, Tasca and Smith–Littau are all appropriate to addressing interconnects. The distinction is the boundary conditions and geometric distinctions of modern interconnect systems. Instead of repeating an exhaustive list of interconnect models, new issues such as Ti/Al/Ti, Cu, low-k dielectrics, fill shapes, voiding, ESD electromigration studies and other issues are briefly addressed. These models are also suitable and applicable for understanding ESD failures in magnetic recording devices such as magnetoresisitive heads, giant magnetoresistive heads, and tunneling magnetoresistive heads.

High-current robustness of interconnects was a concern in the semiconductor industry because of the need to evaluate component survival in military applications[1–9]. Electrothermal models were developed to evaluate the power to failure of the interconnects during an electromagnetic pulse (EMP) event. The focus historically was to provide an understanding as a function of a broadband pulse. The models of Wunsch–Bell [1], Tasca [2,3], Smith [4–6], Egelkrout [7], Kinsbron [8] and Pierce [9] addressed the reliability of the interconnects as a function of pulse width and the resistor geometry for passivated and unpassivated elements.

In today's advanced technologies, the ESD issue is scaling [10–12]. With the continued improvement of MOSFET performance, and an increase in circuit density with each successive technology generation, on-chip interconnect wiring and via scaling must keep pace with the MOSFET constant electric field scaling. With the decreasing

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

transistor gate delay, the on-chip interconnect RC delay contribution plays a more significant role in receiver delay [13–20].

Additionally, as the total number of logic gates increases, the number of external I/O connections also increase. This relationship is known as Rent's rule

$$N_{\mathrm{I/O}} = A\left(N_{\mathrm{gates}}\right)^{B}$$

where $N_{\mathrm{I/O}}$ is the number of peripheral input-output (I/O) circuits, N_{gates} is the number of logic gates, and the parameters A and B are constants, dependent on the technology, type, and application of the semiconductor chips. It has been shown that the constant $A = 0.5$ and $B = 0.5$. As a result, as the number of circuits on a chip increase, the number of I/O increases, requiring smaller interconnect width for wireability. This is inherently valid in high-pin-count applications where electrical connections are arranged in an array configuration as opposed to the peripheral of a semiconductor chip.

In today's advanced high-performance technologies, the need for reduced resistance and capacitance interconnects has lead to an evolution from Ti/Al/Ti or Al-based interconnect systems to copper (Cu)-based interconnect systems. To further improve interconnect performance, low-k dielectrics have been introduced.

With the demand for circuit density, wireability, array I/O architecture, and receiver performance requirements, aluminum interconnects have become a dominant ESD failure mechanism. For example, the interconnect delay component of a CMOS gate delay, excluding the intrinsic delay, can be expressed as

$$\left(T_{\mathrm{g}}\right)_{\mathrm{extrinsic}} = f_{\mathrm{g}} R_{\mathrm{d}} C_{\mathrm{w}} L_{\mathrm{w}} + 0.4 R_{\mathrm{w}} C_{\mathrm{w}} L_{\mathrm{w}}{}^{2} + 0.7 R_{\mathrm{w}} C_{\mathrm{r}} L_{\mathrm{w}}$$

where the extrinsic gate delay involves the interconnect-related delay terms, R_{d} and C_{d} are the MOSFET output resistance and capacitance, respectively, C_{r} is the MOSFET receiver switching capacitance, R_{w},C_{w} and L_{w} are the resistance, capacitance, and the line length of the interconnect, and f_{g} is the circuit fan-out. As the technology is scaled, the interconnect delay terms become an increasingly larger component of the total extrinsic gate delay expression. As a result, the scaling of the interconnects with the MOSFET constant electric field scaling at the interconnect – transistor integration plane. The issue of the scaling of the interconnect system with the technology generation (Figure 9.1) is key to the reason why interconnect ESD robustness is an issue in advanced technologies.

The transition from Al-based to Cu-based interconnect systems, and the transition from silicon dioxide-based inter-level dielectric films to low-k dielectric films has a significant effect on the ESD robustness of the interconnect system. The material change influences the critical current to failure as well as the failure mechanisms.

9.1 ALUMINUM INTERCONNECTS

High-current robustness of interconnects was a concern in the semiconductor industry because of the need to evaluate component survival in military applications. Aluminum interconnects became a standard in semiconductors because of the compatibility with

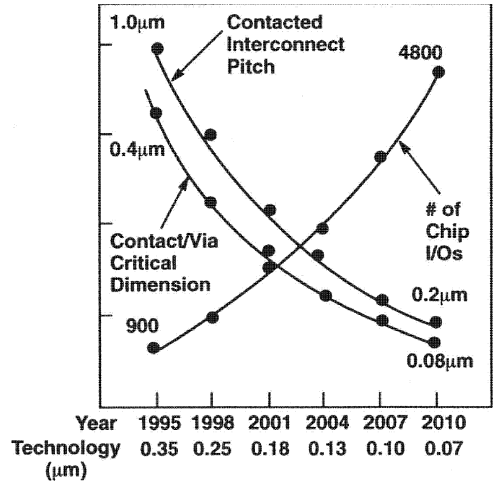

Figure 9.1 Interconnect and via scaling as a function of technology generation

the inter-level dielectric, material properties and the ability to define the interconnects with masking techniques.

Aluminum interconnects are formed on the planar dielectric film. Aluminum interconnects consist of a refractory metal film below and above the aluminum film. The refractory metal assists in the adhesion of the aluminum with the interlevel dielectric material. Typically, a standard process consists of titanium (Ti) and titanium nitride (TiN) where the TiN provides the adhesion between the dielectric and Ti film. To quantify the ESD robustness of a Ti/Al/Ti interconnect, the geometric definition is important for its evaluation. Hence a model must be established to address both the thermal physics and geometry [21].

Resistance is defined according to the linear relationship

$$R(T) = R_0(1 + \alpha T)$$

where $R(T)$ is the dynamic resistance at the temperature T, R_0 is the initial resistance, and α is the temperature coefficient of resistance (TCR). From this form, we can solve for temperature in the interconnect, as

$$T = \frac{R(T) - R_0}{\alpha R_0} = \frac{1}{\alpha}\left(\frac{\Delta R}{R_0}\right)$$

To relate power and thermal impedance to the resistance, the differential resistance $\mathrm{d}R(T)$ is expressed as

$$\mathrm{d}R = \alpha R_0 \mathrm{d}T$$

where we define temperature as the product of the heat fluence and the thermal impedance

$$T = q\theta_{\mathrm{TH}}$$

The total differential of temperature can be expressed as

$$dT = q\theta_{TH} + \theta_{TH}dq$$

The heat flux is equal to the input power from conservation of energy. Substituting in terms of power for the heat flux, the total differential resistance can be expressed as

$$\frac{dR}{R_0} = \alpha(Pd\theta + \theta dP)$$

Since the impedance is constant, and integration of the expression from the initial to final resistance

$$\frac{\Delta R}{R} = \frac{R_f - R_0}{R_0} = \alpha\theta_{TH}P$$

Hence, the normalized resistance change is proportional to the product of the thermal impedance, power and temperature coefficient of resistance. Combining Joule heating ($P = I^2R$) and resistance ($R = \rho L/A$), and substituting in for the current density J, the the expression of normalized differential resistance can be expressed as [21]

$$\frac{\Delta R}{R} = \alpha\theta_{TH}J^2\rho LA$$

From the normalized differential resistance, we can solve for the thermal impedance as,

$$\theta_{TH} = \frac{1}{J^2}\frac{\lfloor \Delta R/R \rfloor}{\alpha\rho LA}$$

From this formulation, the thermal impedance can be extracted from the resistance change and the current density.

In a Ti/Al/Ti interconnect, the interconnect cross-section is rectangular. The liner film exists on the top and the bottom of the film, but not on the sides of the interconnect (Figure 9.2). Since the interconnect is a composite film, with the refractory metal having

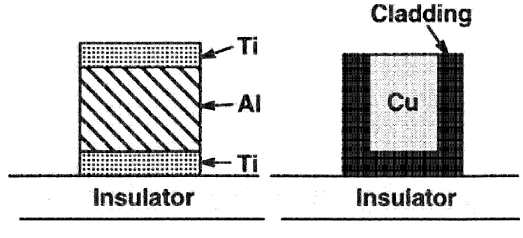

Figure 9.2 Aluminum and copper interconnect structures

a high resistance value, an effective resistivity model can be used where we assume the effective resistivity is

$$\rho_{\text{eff}} = \frac{\rho_{\text{Al}}}{(1 - 2\delta/y)} = \frac{\rho_{\text{Al}}}{\aleph}$$

where the geometrical dimensions are the liner bottom δ and line height y.

Assuming a critical energy E_{crit}, at which the interconnect failure occurs, this can be related to the product of the average energy input during a pulse event. Assuming that the current is constant during the pulse, time, and the current can be removed from the average, where the average resistance is the the mean between the initial and final resistance value.

$$E_{\text{crit}} = \int_0^{\Delta t} P(t)\mathrm{d}t = I^2 \int_0^{\Delta t} R(t)\mathrm{d}t = \frac{1}{2}[R_0 + R_{\text{f}}]I^2\Delta t$$

We can define an expression which is the normalized change in resistance as

$$\gamma_{\text{crit}} = \frac{(R_{\text{f}} - R_0)}{R_0}$$

and express the critical energy according to the normalized change in resistance at failure

$$E_{\text{crit}} = \frac{1}{2}[R_0(2 + \gamma_{\text{crit}})]I^2\Delta t$$

Energy can be related to the heat capacity and heat of fusion,

$$E = C\Delta T + m_i L_{\text{f}}$$

Hence, when the temperature is the critical temperature of failure, then the energy expression is the critical energy. From the resistance relationship with temperature, we can substitute in for the temperature change as a function of normalized resistance change, and the temperature coefficient of resistance

$$E_{\text{crit}} = C\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}$$

In the case of the composite film, the mass of each region, and specific heat of each film must be defined. The above development implies that the temperature of the composite film is the same. More accurately, we can find the total mass summed over all regions

$$E_{\text{crit}} = \sum m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}$$

From the two forms of the critical energy

$$I_{\text{crit}}{}^2 = \frac{\sum_i m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}}{\frac{1}{2}\Delta t R_0 (2 + \gamma_{\text{crit}})}$$

Expanding this expression to address the multiple films (e.g. aluminum, cladding and insulator materials)

$$I_{\text{crit}}{}^2 = \frac{\left[\left(m_i c_i + m_j c_j\right)\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}} + m_k c_k \left(\frac{\gamma_{\text{crit}}}{\alpha}\right)\right]}{\frac{1}{2}\Delta t\left(\rho_{\text{eff}} \frac{L}{A}\right)(2 + \gamma_{\text{crit}})}$$

In this expression, the cross-sectional area of the interconnect can be brought into the numerator in order to express the equation as a function of the linewidth and the geometric variables associated with the films.

The physical volume of the Al region can be expressed as

$$V_{\text{Al}} = LWd_{\text{s}}\aleph$$

and

$$\aleph = \left(1 - \frac{2\delta}{d_{\text{s}}}\right)$$

The cladding volume can be expressed as

$$V_{\text{Clad}} = LWd_{\text{s}} - LWd_{\text{s}}\aleph = LWd_{\text{s}}[1 - \aleph] = LW[2\delta]$$

The fully passivated thermal sheath volume can be expressed as the volume of the interconnect with the additional dimension associated with the thermal heat volume in the insulator minus the interconnect

$$V_{\text{sheath}} = \{L + 2d_{\text{ox}}\}\{W + 2d_{\text{ox}}\}\{d_s + 2d_{\text{ox}}\} - LWd_{\text{s}}$$

This form assumes extension beyond the length of the interconnect. This is in the form of

$$V_{\text{sheath}} = 4[W + d_{\text{s}} + L]d_{\text{ox}}{}^2 + 2[LW + Ld_{\text{s}} + 2Wd_{\text{s}}]d_{\text{ox}}$$

and the time-dependent oxide sheath thickness is a function of the pulse width, and the heat diffusion coefficient in the oxide

$$d_{\text{ox}}(t) = a_{\text{d}}\sqrt{\Delta t}$$

With the assumption, that the highest level film is unpassivated, this can be expressed as

$$V_{\text{sheath}} = \{L + 2d_{\text{ox}}\}\{W + 2d_{\text{ox}}\}\{d_{\text{s}} + d_{\text{ox}}\} - LWd_{\text{s}}$$

As an additional assumption, for long structures with passivation above and below the interconnect, we can ignore the thermal component on the end of the wires and simplify the analysis to the case of

$$V_{\text{sheath}} = \{L\}\{W + 2d_{\text{ox}}\}\{d_{\text{s}} + 2d_{\text{ox}}\} - LWd_{\text{s}}$$

Let the region be such that the indices i,j,k be in the order of Al, cladding and the insulator volume, respectively.

Expressing this term as a function of the volumes, heat capacity and density

$$I_{\text{crit}}{}^2 = \frac{2Wd_{\text{s}}\left[\left(\psi_i V_i C_i + \psi_j V_j C_j\right)\left(\frac{\gamma_{crit}}{\alpha}\right) + \psi_i V_i L_{\text{f}} + \psi_k V_k c_k \left(\frac{\gamma_{\text{crit}}}{\alpha}\right)\right]}{\Delta t(\rho_{\text{eff}} L)(2 + \gamma_{\text{crit}})}$$

ESD robustness of aluminum interconnects has been evaluated using HBM, MM and TLP measurement as a function of design width. Figure 9.3 shows a plot of the HBM results of aluminum interconnects as a function of design width. As the linewidth increases, the ESD robustness also increases in a linear fashion.

Figure 9.4 shows the dynamic resistance change normalized to the initial resistance as a function of the current density. The plot of $\Delta R/R_0$-J for a Ti/Al/Ti interconnect is a function of pulse width.

Figure 9.5 is the permanent d.c. resistance shift observed in the interconnect post-ESD stress.

From the experimental results on failure, the critical current density to failure can be plotted as a function of the applied pulse width. Figure 9.6 shows the plot for M1 and M3 interconnect wires. As the wire moves farther from the silicon surface, the thermal resistance to the bulk increases, decreasing the ESD robustness of the interconnect.

Figure 9.7 is a cross-section drawing of an Ti/Al/Ti interconnect during ESD failure. Typically, the aluminum melts from the self-heating. The center of the interconnect is

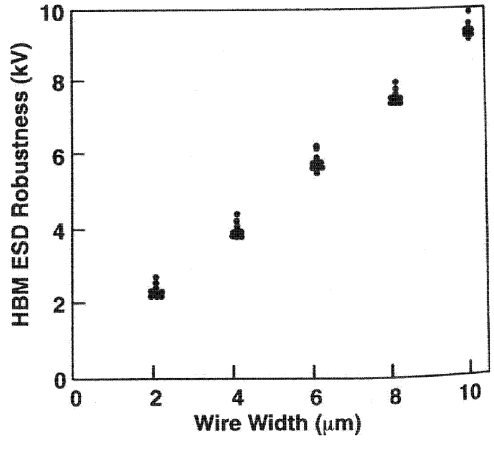

Figure 9.3 HBM ESD robustness of an aluminum interconnect as a function of wire width (M4)

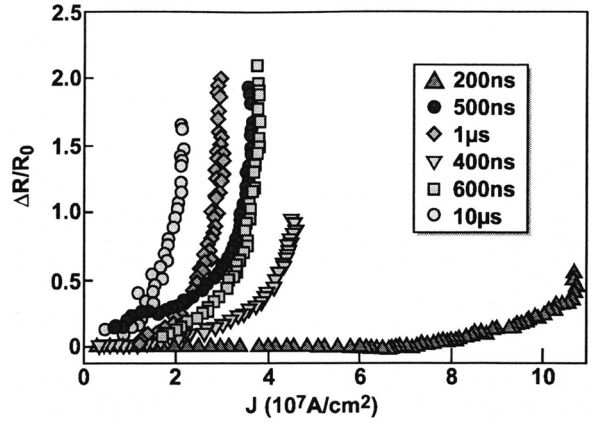

Figure 9.4 Dynamic $\Delta R/R_0$ as a function of J for a Ti/Al/Ti interconnect for different values of pulse width

the location of peak temperature. As the heating occurs, the insulator cracks laterally, leading to the liquid aluminum entering the insulator region. As the wire interconnect continues to be pulsed, eventually the Ti cladding material begins to melt.

Figure 9.8 is an example of a contact failure in the area of a MOSFET structure. From the Wrights' etch pattern, it can be observed that the region of peak heating follow the relationship of the square of the electric field pattern.

In Ti/Al/Ti interconnect systems, the vias are made of tungsten studs. Tungsten has a high melting temperature, well above those of the other materials in semiconductor devices. Figures 9.9 and 9.10 are the HBM and MM results of a tungsten stud via structure as a function of the number of vias.

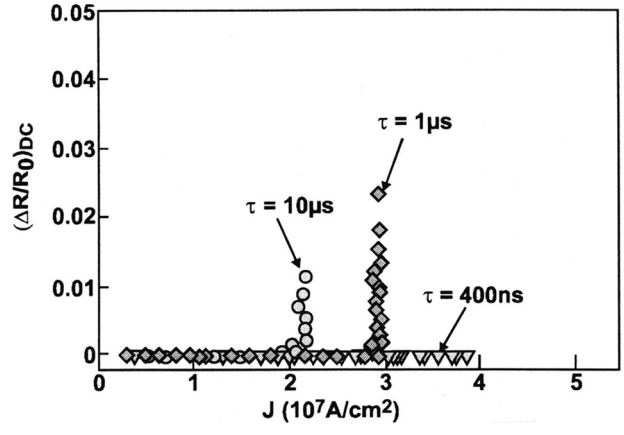

Figure 9.5 Normalized d.c. resistance $\Delta R/R_0$ as a function of J for a Ti/Al/Ti interconnect

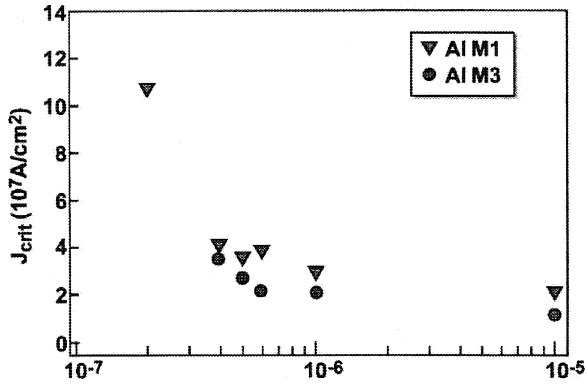

Figure 9.6 Critical current density to failure as a function of interconnect pulse width τ (ns) for M1 and M3 wires

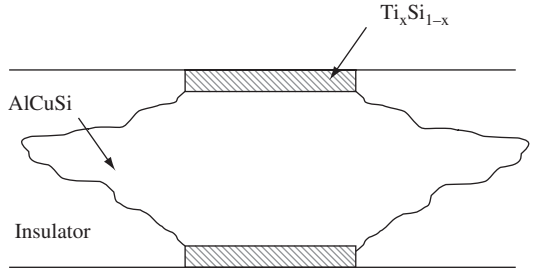

Figure 9.7 Cross-section of an aluminum wire after ESD failure

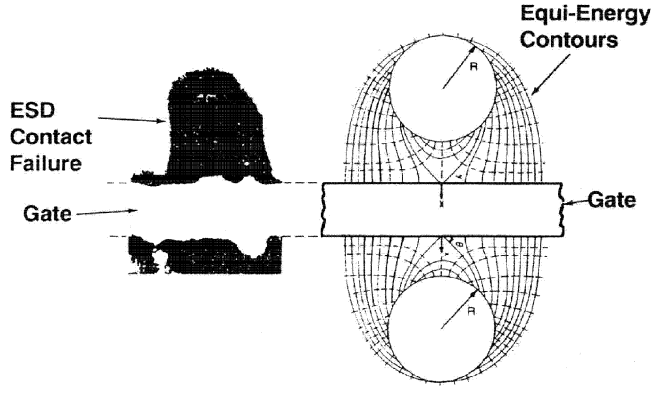

Figure 9.8 ESD damage in contact area in a MOSFET structure

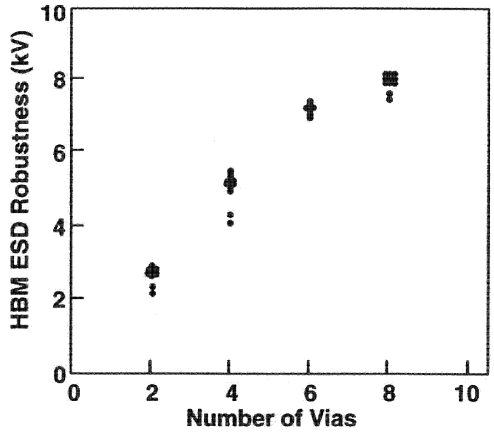

Figure 9.9 HBM ESD robustness of a tungsten via as a function of number of vias (e.g. in parallel)

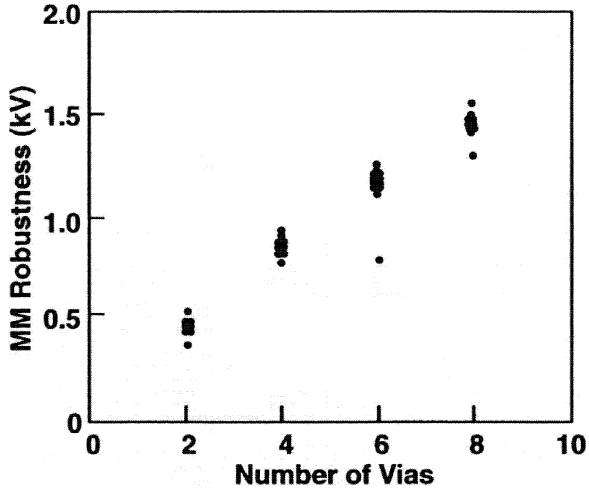

Figure 9.10 MM ESD robustness of a tungsten via as a function of number of vias (e.g. in parallel)

9.2 COPPER INTERCONNECTS

In today's advanced high-performance technologies, the need for reduced resistance and capacitance interconnects has led to an evolution from Ti/Al/Ti or Al-based interconnect systems to copper (Cu)-based interconnect systems. Copper interconnect systems (Figure 9.11) are formed by providing troughs in the inter-level dielectric (ILD) films. The dielectric is etched using reactive ion etch (RIE) processes, followed by cladding material and Cu film deposition. The cladding, or liner material, is typically a refractory metal film. The cladding serves as a diffusion barrier, and provides adhesion to the insulator film. These materials can include TiN, WN, Ta, TaN, or TaSiN [22–27].

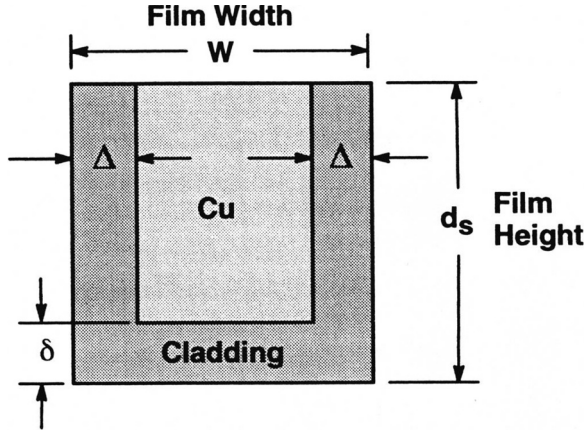

Figure 9.11 Cross-section of a Cu interconnect

To quantify the ESD robustness of a copper interconnect, the material and geometry is required. Copper interconnects in the damascene and dual-damascene provide both different failure levels and mechanisms compared to an aluminum-based interconnect system. Hence a model must be established to address both the thermal physics and geometry.

As in the case of the aluminum interconnect, resistance is defined according to the linear relationship

$$R(T) = R_0(1 + \alpha T)$$

where $R(T)$ is the dynamic resistance at the temperature T, R_0 is the initial resistance, and α is the temperature coefficient of resistance (TCR). From this form, we can solve for temperature in the interconnect, as

$$T = \frac{R(T) - R_0}{\alpha R_0} = \frac{1}{\alpha}\left(\frac{\Delta R}{R_0}\right)$$

To relate power and thermal impedance to the resistance, the differential resistance $\mathrm{d}R(T)$ is expressed as

$$\mathrm{d}R = \alpha R_0 \mathrm{d}T$$

where we define temperature as the product of the heat flux and the thermal impedance

$$T = q\theta_{\mathrm{TH}}$$

The total differential of temperature can be expressed as

$$\mathrm{d}T = q\theta_{\mathrm{TH}} + \theta_{\mathrm{TH}}\mathrm{d}q$$

The heat flux is equal to the input power, from conservation of energy. Substituting in terms of power for the heat flux, the total differential resistance can be expressed as

$$\frac{\mathrm{d}R}{R_0} = \alpha(P\mathrm{d}\theta + \theta\mathrm{d}P)$$

Since the impedance is constant, and integration of the expression from the initial to final resistance

$$\frac{\Delta R}{R} = \frac{R_\mathrm{f} - R_0}{R_0} = \alpha\theta_\mathrm{TH}P$$

Hence, the normalized resistance change is proportional to the product of the thermal impedance, power and temperature coefficient of resistance. Combining Joule heating ($P = I^2R$) and resistance ($R = \rho L/A$), and substituting in for the current density J, then the expression of normalized differential resistance can be expressed as

$$\frac{\Delta R}{R} = \alpha\theta_\mathrm{TH}J^2\rho LA$$

From the normalized differential resistance, we can solve for the thermal impedance as

$$\theta_\mathrm{TH} = \frac{1}{J^2}\frac{\lfloor \Delta R/R \rfloor}{\alpha\rho LA}$$

From this formulation, the thermal impedance can be extracted from the resistance change and the current density.

Figure 8.3 shows the geometry of the damascene copper interconnect. In a damascene copper interconnect, the interconnect cross-section is rectangular.

The liner film exists on three sidewalls, but not on the top region. Since the interconnect is a composite film, with the refractory metal being of high resistance, an effective resistivity model can be used where we assume the effective resistivity is [28–30].

$$\rho_\mathrm{eff} = \frac{\rho_\mathrm{Cu}}{(1 - 2\Delta/x)(1 - 2\delta/y)} = \frac{\rho_\mathrm{Cu}}{\aleph}$$

where the geometric dimensions are the liner sidewall thickness Δ, liner bottom δ, linewidth x, and line height y.

Assuming a critical energy E_crit, at which the interconnect failure occurs, this can be related to the product of the average energy input during a pulse event. Assuming that the current is constant during the pulse, time and current can be removed from the average, where the average resistance is the the mean between the initial and final resistance values.

$$E_\mathrm{crit} = \int_0^{\Delta t} P(t)\mathrm{d}t = I^2\int_0^{\Delta t} R(t)\mathrm{d}t = \frac{1}{2}[R_0 + R_\mathrm{f}]I^2\Delta t$$

We can define an expression which is the normalized change in resistance as

$$\gamma_{\text{crit}} = \frac{(R_{\text{f}} - R_0)}{R_0}$$

and express the critical energy according to the normalized change in resistance at failure

$$E_{\text{crit}} = \frac{1}{2}[R_0(2 + \gamma_{\text{crit}})]I^2 \Delta t$$

Energy can be related to the heat capacity and heat of fusion

$$E = C\Delta T + m_i L_{\text{f}}$$

Hence, when the temperature is the critical temperature of failure, then the energy expression is the critical energy. From the resistance relationship to temperature, we can substitute for the temperature change as a function of normalized resistance change, and the temperature coefficient of resistance

$$E_{\text{crit}} = C\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}$$

In the case of the composite film, the mass of each region, and specific heat of each film must be defined. In the above development, it implies that the temperature of the composite film is the same. More accurately, we can find the total mass summed over all regions

$$E_{\text{crit}} = \sum m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}$$

From the two forms of the critical energy

$$I_{\text{crit}}{}^2 = \frac{\sum\limits_i m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}}{\frac{1}{2}\Delta t R_0 (2 + \gamma_{\text{crit}})}$$

Expanding this expression to address the multiple films (e.g. copper, cladding and insulator materials)

$$I_{\text{crit}}{}^2 = \frac{\left[\left(m_i c_i + m_j c_j\right)\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}} + m_k c_k \left(\frac{\gamma_{\text{crit}}}{\alpha}\right)\right]}{\frac{1}{2}\Delta t\left(\rho_{\text{eff}} \frac{L}{A}\right)(2 + \gamma_{\text{crit}})}$$

In this expression, the cross-sectional area of the interconnect can be brought into the numerator in order to express the equation as a function of the linewidth and the geometric variables associated with the films.

The physical volume of the Cu region can be expressed as [28,29]

$$V_{\mathrm{Cu}} = LWd_s\aleph$$

and

$$\aleph = \left(1 - \frac{2\Delta}{W}\right)\left(1 - \frac{2\delta}{d_s}\right)$$

The cladding volume can be expressed as

$$V_{\mathrm{Clad}} = L\left[W\left(1 - \frac{2\Delta}{W}\right)\delta + 2d_s\Delta\right]$$

The fully passivated thermal sheath volume can be expressed as

$$V_{\mathrm{sheath}} = 4[W + d_s + L]{d_{\mathrm{ox}}}^2 + 2[LW + Ld_s + 2Wd_s]d_{\mathrm{ox}}$$

and the time-dependent oxide sheath thickness is a function of the pulse width, and the heat diffusion coefficient in the oxide

$$d_{\mathrm{ox}}(t) = a_{\mathrm{d}}\sqrt{\Delta t}$$

Let the region be such that the indices i, j, k be in the order of Cu, cladding and the insulator volume, respectively.

Expressing this term as a function of the volumes, heat capacity and density

$${I_{\mathrm{crit}}}^2 = \frac{2Wd_s\left[\left(\psi_i V_i C_i + \psi_j V_j C_j\right)\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right) + \psi_i V_i L_{\mathrm{f}} + \psi_k V_k c_k\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right)\right]}{\Delta t(\rho_{\mathrm{eff}}L)(2 + \gamma_{\mathrm{crit}})}$$

Figure 9.12 is a plot of the HBM ESD robustness of a Cu interconnect as a function of the wire width. From the experimental results, the ESD robustness increases linearly with the width of the wire interconnect.

From transmission line pulse testing, the resistance shift as a function of current density can be plotted as a function of pulse width. Figures 9.13 and 9.14 show the change in the wire resistance as a function of the current and the square of the current density.

Figure 9.15 is the Wunsch–Bell plot of the critical current density to failure as a function of pulse width for a Cu interconnect.

Figure 9.16 is the critical current density to failure comparison of copper and aluminum interconnects. Because of the higher melting temperature, Cu interconnect failure occurs at approximately double the Al interconnect failure level.

Evaluation of the Cu interconnect after ESD stress shows damage patterns forming in the center of the linewidth. Figure 9.17 is an example of the Cu interconnect after ESD stress.

In passivated interconnects, after the Cu interconnect melting, the insulator cracks above the interconnect. The insulator crack propagates upward from the edges of the

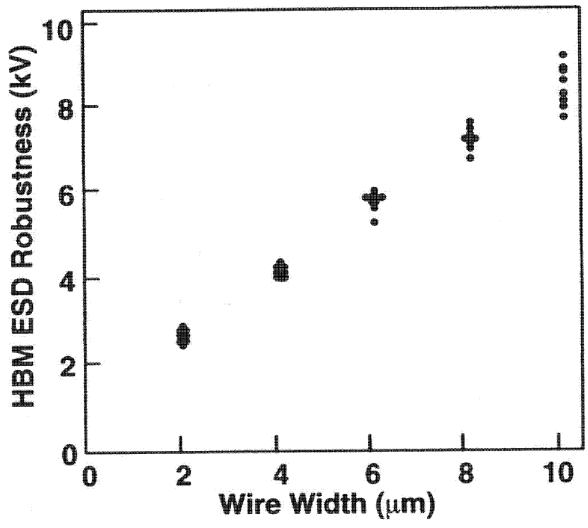

Figure 9.12 HBM ESD robustness of a copper interconnect as a function of wire width (M3)

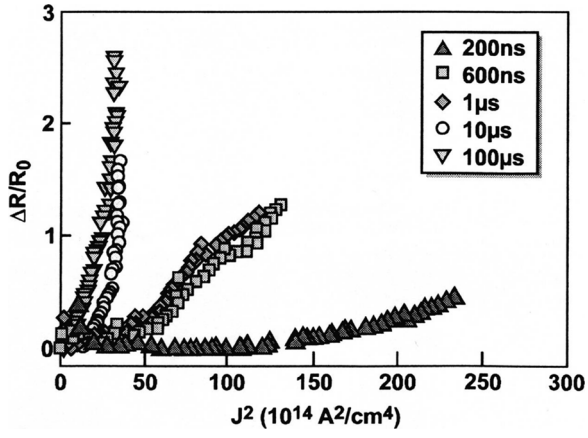

Figure 9.13 Dynamic $\Delta R/R_0$ as a function of J for a copper interconnect

interconnect. Figure 9.18 is a representation of the cracking damage pattern of the Cu interconnect.

9.2.1 Copper Vias

The ESD robustness of copper vias are of interest for the design of electrical interconnects to insure the interconnect integrity in the interconnect and insulator materials.

In a Cu-based interconnect system, the vias are formed as part of the same physical film deposition. A first and second trough are formed using a first and second reactive ion etch (RIE) process, followed by the liner deposition, and then the copper deposition.

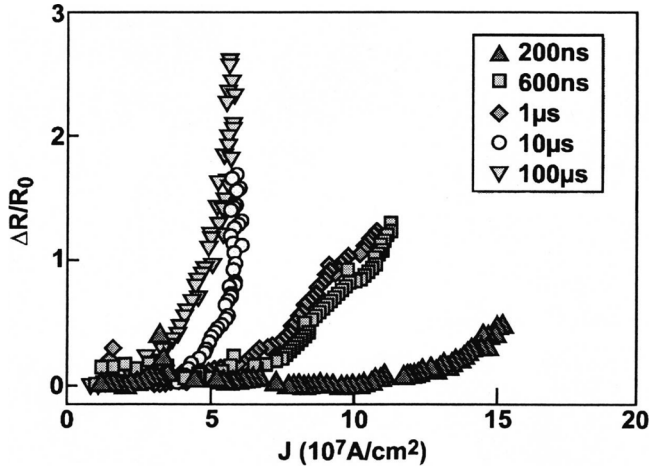

Figure 9.14 Dynamic $\Delta R/R_0$ as a function of J^2 for a copper interconnect

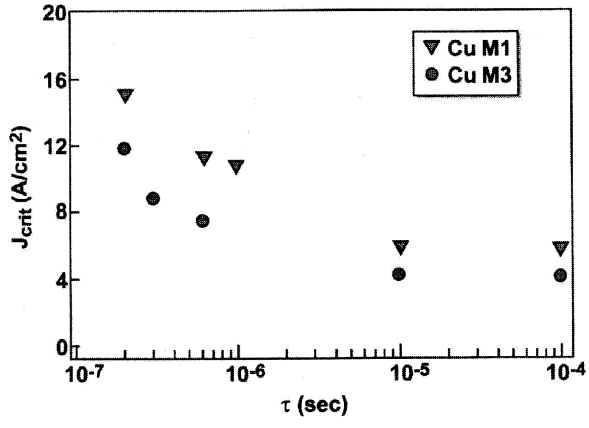

Figure 9.15 Critical current density to failure as a function of pulse width for Cu interconnects

The Cu interconnect/via structure is completed by a chemical mechanical polishing to planarize the structure.

In a dual-damascene process, the Cu via volume can be expressed as[29]

$$V_{\mathrm{Cu}} = ht^2\left(1-\frac{2\Delta}{t}\right)^2\left(1-\frac{\delta}{h}\right)$$

where the via height is h, the via edge is t, with a cladding bottom thickness δ, and sidewall film thickness Δ.

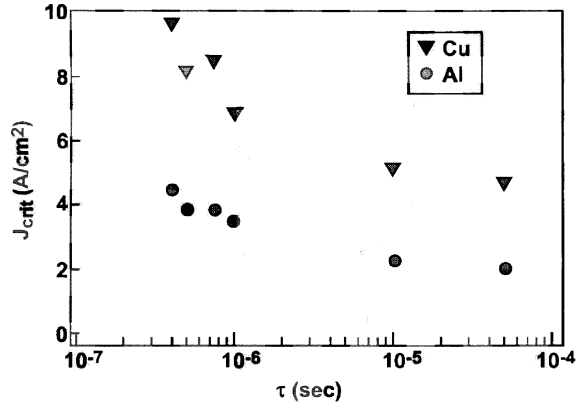

Figure 9.16 Critical current density to failure comparison of copper and aluminum interconnects

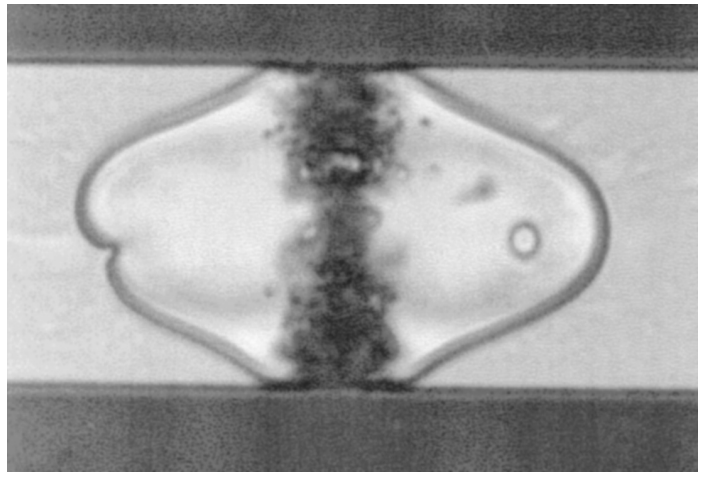

Figure 9.17 ESD failure of a copper interconnect

The volume of the cladding material includes the bottom film, and the sidewalls are the outer dimension of the structure without the Cu film inside.

$$V_{\mathrm{Cu}} = ht^2\left[1 - \left(1 - \frac{2\Delta}{t}\right)^2\left(1 - \frac{\delta}{h}\right)\right]$$

For ESD pulse events where the heat diffuses into the insulator, a thermal sheath is formed around the interconnect. The thickness of this sheath can be referred to as d_{ox}. Hence the volume in the insulator

$$V_{\mathrm{sheath}} = ht^2\left[\left(1 + \frac{2d_{\mathrm{ox}}}{t}\right)^2 - 1\right]$$

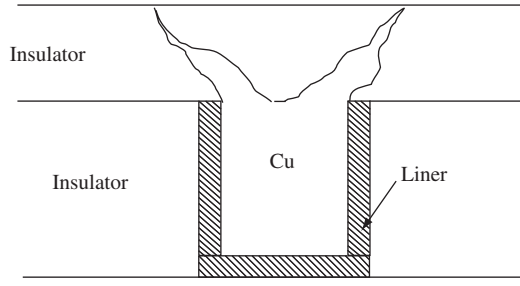

Figure 9.18 Cu interconnect cross-section, highlighting the insulator cracking from thermal stress post-ESD testing

and the time-dependent oxide sheath thickness is a function of the pulse width, and the heat diffusion coefficient in the oxide

$$d_{\mathrm{ox}}(t) = a_{\mathrm{d}}\sqrt{\Delta t}$$

Solving for the critical current to failure of the via structure, the expression for the critical current as a function of the multiple films (e.g. copper, cladding and insulator materials) can be applied

$$I_{\mathrm{crit}}{}^2 = \frac{\left[\left(m_i c_i + m_j c_j\right)\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right) + m_i L_{\mathrm{f}} + m_k c_k\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right)\right]}{\frac{1}{2}\Delta t\left(\rho_{\mathrm{eff}}\frac{L}{A}\right)(2+\gamma_{\mathrm{crit}})}$$

Let the region be such that the indices i, j, k be in the order of Cu via, via cladding and the insulator volume around the via structure, respectively.

Expressing this term as a function of the volumes, heat capacity and density

$$I_{\mathrm{crit}}{}^2 = \frac{2ht\left[\left(\psi_i V_i C_i + \psi_j V_j C_j\right)\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right) + \psi_i V_i L_{\mathrm{f}} + \psi_k V_k c_k\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right)\right]}{\Delta t(\rho_{\mathrm{eff}}\, t)(2+\gamma_{\mathrm{crit}})}$$

In this expression, the effective resistivity is associated with that of the via structure. This is different compared to the formula used for the linewidth. In this case we can express the effective resistivity as

$$\rho_{\mathrm{eff}} = \frac{\rho_{\mathrm{Cu}}}{\left(1-\frac{2\Delta}{t}\right)^2} = \frac{\rho_{\mathrm{Cu}}}{\aleph_{\mathrm{via}}}$$

Figure 9.19 is a plot of the critical current to failure as a function of pulse width for a Cu via. As the pulse width increases, the critical current to failure decreases.

In Figure 9.20, an ESD comparison of tungsten (W) stud and copper dual-damascene vias is shown. From these results, the Cu damascene via is superior to the W stud via. Because of the damascene single film nature, the metal line also participates in the self-heating.

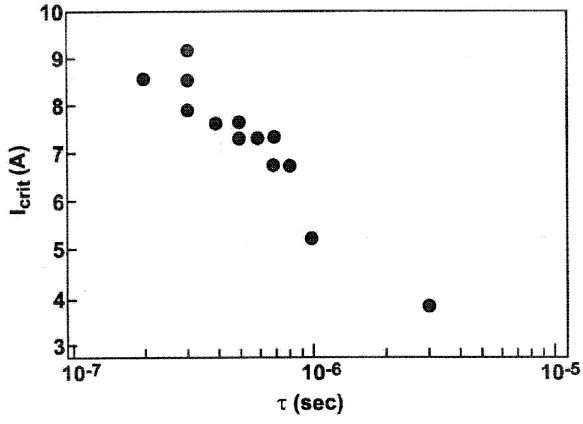

Figure 9.19 Copper via critical curent to failure as a function of pulse width

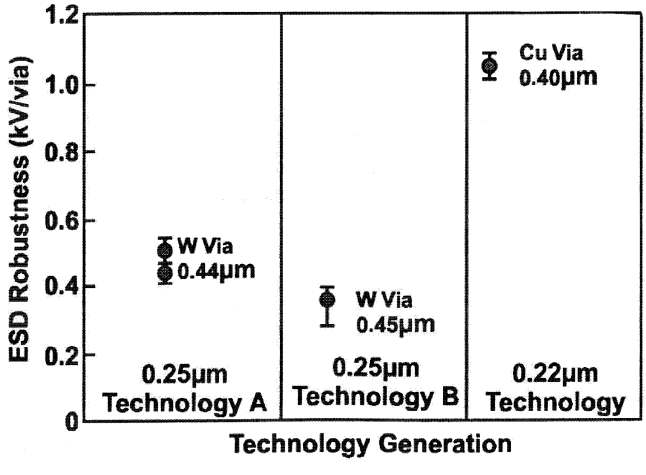

Figure 9.20 Comparison of tungsten and copper via

9.3 LOW-κ MATERIALS AND INTERCONNECTS

With the demand for circuit density, wireability, array I/O architecture, and receiver performance requirements, low-k dielectric as interlevel dielectric films have become important in signal delay [31–37]. For example, the interconnect delay component of a CMOS gate delay, excluding the intrinsic delay, can be expressed as

$$(T_g)_{\text{extrinsic}} = f_g R_d C_w L_w + 0.4 R_w C_w L_w^2 + 0.7 R_w C_r L_w$$

where the extrinsic gate delay are the interconnect-related delay terms, R_d and C_d are the MOSFET output resistance and capacitance, respectively, C_r is the MOSFET receiver switching capacitance, R_w, C_w, and L_w are the resistance, capacitance, and the line length of the interconnect, and f_g is the circuit fan-out. To reduce the interconnect delay

time, lowering the capacitance of the interconnect plays a significant component. By using copper interconnects, the resistance of the interconnect can be reduced, but the low-k ILD film allows for reduction of the interconnect capacitance and the line-to-line coupling. By using copper interconnects and low-k ILD films, the signal-delay characteristics can be met to achieve continued performance with technology scaling.

The scaling of the dielectric permittivity with MOSFET scaling is achieved if scaling follows

$$k' = \frac{k}{\alpha}$$

where they are scaled according to the MOSFET constant electric field scaling parameter. Low-k interlevel dielectric films must also have low leakage, compatible thermal coefficients of expansion, high breakdown voltage, low film stress, high cracking tolerance, and compatibility with the interconnect systems (e.g. damascene Cu interconnects). Typical low-k materials include silicon oxyfluoride $(F_xSi)_y$, hydrogen silsequioxane (HSQ), organic polymers, such as poly(arelene) ethers, benzocyclobutene (BCB), nanoporous silica and porous polymers.

Figure 9.21 shows the proposed technology roadmap for interconnects and low-k material. This technology roadmap shows the scaling of the interconnect width, the increasing number of I/O networks, and the scaling of the interconnect material.

ESD measurements on aluminum interconnects in a low-k gap fill material demonstrated that the ESD robustness of an aluminum interconnect will decrease compared with a silicon dioxide interlevel dielectric [38].

From an ESD perspective, the thermal characteristics of the low-k material will influence the thermal diffusion time τ_D, the thermal diffusivity $\alpha_{>D}$, and the Fourier number

$$F_0 = \frac{K}{\rho C_p}\frac{t}{L^2}$$

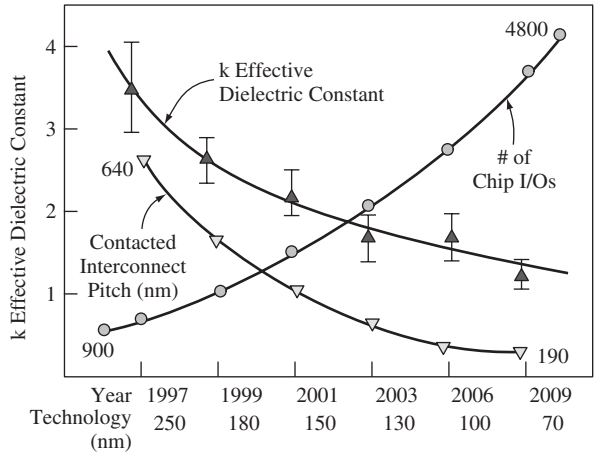

Figure 9.21 Low-k interlevel dielectric (ILD) electrical permittivity as a function of technology generation and year (Source: SIA National Technology Roadmap for Semiconductors)

where t and L are the characteristic time and length scale. For a damascene interconnect in a low-k ILD maternal, the physical volume of the Cu region can be expressed as [39]

$$V_{\mathrm{Cu}} = LWd_{\mathrm{s}}\aleph$$

and

$$\aleph = \left(1 - \frac{2\Delta}{W}\right)\left(1 - \frac{2\delta}{d_{\mathrm{s}}}\right)$$

The cladding volume can be expressed as

$$V_{\mathrm{Clad}} = L\left[W\left(1 - \frac{2\Delta}{W}\right)\delta + 2d_{\mathrm{s}}\Delta\right]$$

The fully passivated thermal sheath volume can be expressed as

$$V_{\mathrm{sheath}} = 4[W + d_{\mathrm{s}} + L]{d_{\mathrm{ILD}}}^2 + 2[LW + Ld_{\mathrm{s}} + 2Wd_{\mathrm{s}}]d_{\mathrm{ILD}}$$

and the time-dependent low-k ILD sheath thickness is a function of the pulse width

$$d_{\mathrm{ox}}(t) = a_{\mathrm{d}}\sqrt{\Delta t}$$

Let the region be such that the indices i, j, k be in the order of Cu, cladding and the low-k insulator volume, respectively.

Expressing this term as a function of the volumes, heat capacity and density

$${I_{\mathrm{crit}}}^2 = \frac{2Wd_{\mathrm{s}}\left[\left(\psi_i V_i C_i + \psi_j V_j C_j\right)\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right) + \psi_i V_i L_{\mathrm{f}} + \psi_k V_k c_k\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right)\right]}{\Delta t(\rho_{\mathrm{eff}} L)(2 + \gamma_{\mathrm{crit}})}$$

To simplify the analysis, we will assume the length of the interconnect is significantly greater than the width, and the width is significantly greater than the cladding thickness.

$$V_{\mathrm{Cu}} = LWd_{\mathrm{s}}\aleph$$

$$V_{\mathrm{Clad}} \cong LW\delta$$

$$V_{\mathrm{sheath}} \cong 2WLd_{\mathrm{ILD}}$$

and for an unpassivated film

$$V_{\mathrm{sheath}} \cong WLd_{\mathrm{ILD}}$$

From this formulation, the energy can be defined as

$$E = C\Delta T + m_i L_{\mathrm{f}}$$

or as a function of the physical volumes

$$E = [\rho_{Cu} V_{Cu} C_{Cu} + \rho_{Clad} V_{Clad} C_{Clad} + \rho_{ILD} V_{ILD} C_{ILD}] C\Delta T + m_{Cu} L_f$$

Let us define an effective thickness t_{eff}

$$t_{eff} = d_S \left[\aleph + \frac{\rho_{Clad} C_{Clad}}{\rho_{Cu} C_{Cu}} \frac{\delta}{d_S} + \frac{\rho_{ILD} C_{ILD}}{\rho_{Cu} C_{Cu}} \frac{2d_{ILD}}{d_S} \right]$$

Assuming a critical energy E_{crit}, at which the interconnect failure occurs, this can be related to the product of the average energy input during a pulse event. Assuming that the current is constant during the pulse, time, and the current can be removed from the average, where the average resistance is the the mean between the initial and final resistance value.

$$E_{crit} = \int_0^{\Delta t} P(t) \mathrm{d}t = I^2 \int_0^{\Delta t} R(t) \mathrm{d}t = \frac{1}{2} [R_0 + R_f] I^2 \Delta t$$

Solving for the critical current

$$I_{crit}{}^2 = W^2 \left[\frac{2LW[\rho_{Cu} C_{Cu} t_{eff} \Delta T + \rho_{Cu} d_S \aleph L_f]}{R_0 \left(2 + \frac{\Delta R}{R_0} \right) \tau} \right]$$

Substituting in the effective resistivity term, we can express the critical current with the low-k dielectric as [39]

$$I_{crit} = W \sqrt{t t_{eff}} \left[\frac{2 \left[\rho_{Cu} C_{Cu} \Delta T + \rho_{Cu} \left(\frac{t}{t_{eff}} \right) \aleph L_f \right]}{\rho_{eff} \left(2 + \frac{\Delta R}{R_0} \right) \tau} \right]^{1/2}$$

where t is the film thickness of the interconnect ($t = d_S$). From this formulation, the critical current is expressed as the arithmetic mean of the thickness of the interconnect and the effective thickness of the interconnect.

Figures 9.22 and 9.23 are plots of the HBM and MM results comparing the Cu interconnect in silicon dioxide and low-k ILD materials. HBM results show the Cu interconnect in the low-k dielectric has a lower ESD robustness level compared with the silicon dioxide ILD material.

From transmission line pulse testing, the change in the dynamic resistance can be evaluated as a function of the current density. Figures 9.24 and 9.25 show $\Delta R/R_0$–J for an M2 Cu/SiO_2 ILD and low-k ILD interconnect system, respectively.

Figure 9.26 compares critical current to failure for the Cu interconnect in silicon dioxide with the value for low-k ILD dielectrics.

Figure 9.27 shows a Wunsch–Bell J_{crit} curve for an aluminum interconnect in silicon dioxide, a damascene Cu interconnect in silicon dioxide, and a Cu interconnect in

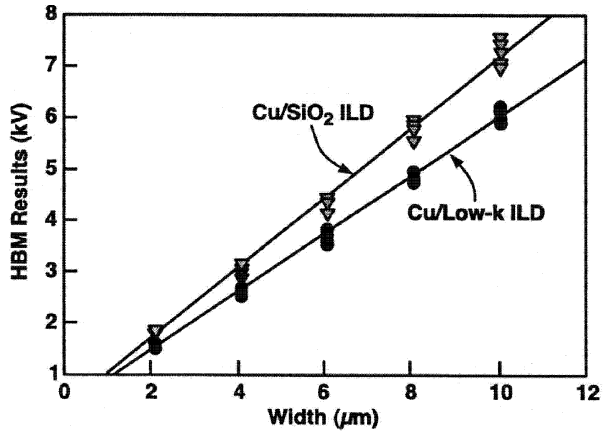

Figure 9.22 HBM comparison of Cu interconnect in silicon dioxide and low-k ILD materials

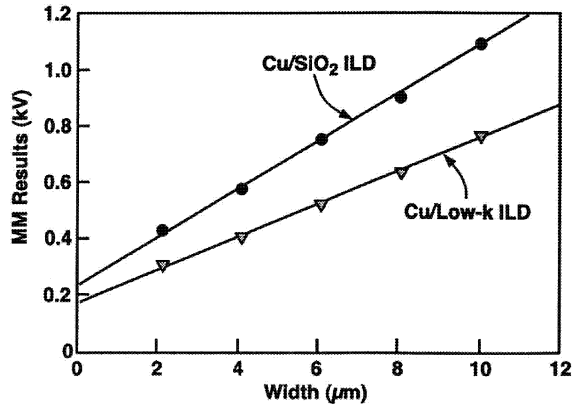

Figure 9.23 MM Comparison of cu interconnect in silicon dioxide and low-k ILD materials

a low-k material. From the experimental results, the ESD degradation associated with low-k material can be observed.

From these results, the degradation due to the low-k material can be recovered by using a wider linewidth for interconnects. Figure 9.28 demonstrates design curves to define the linewidth to achieve a given ESD level for different interconnect and ILD systems.

9.4 POLISHING STOPS AND INTERCONNECTS

For mechanical polishing to avoid dishing effects, insulators are placed in the electrical interconnect after a given metal linewidth. In a typical metal line of width W, these shapes are placed in the metal line.

The placement of insulators in a metal line is an issue for functionality and electrostatic discharge (ESD) robustness. Hence, the model for metal lines must address these openings in the interconnect for wide interconnects.

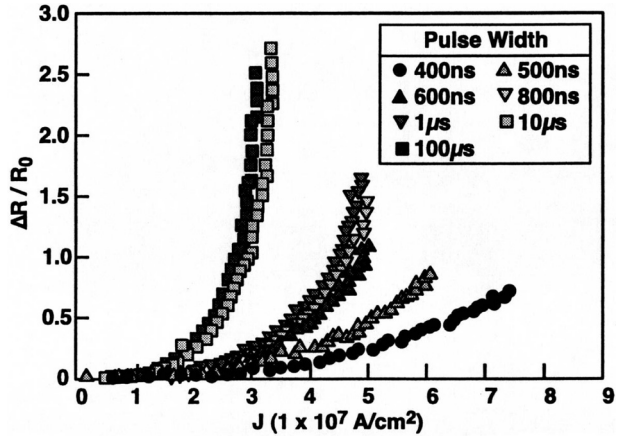

Figure 9.24 $\Delta R/R_0$ as a function of J for an M2 Cu/SiO_2 ILD interconnect system

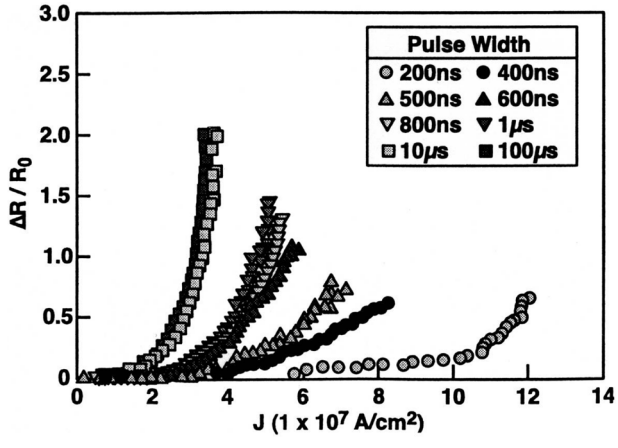

Figure 9.25 $\Delta R/R_0$ as a function of J for a M2 Cu/Low-k ILD interconnect system

Given an interconnect of initial resistance R_0, width W, and length L, film thickness t, and resistivity ρ, we can define a hole in the interconnect of width W' and length L'. The resistance for a length L' where there is one hole every distance L' is

$$R' = \rho\frac{L'-W'}{Wt} + \rho\frac{W'}{(W-W')t} = \frac{\rho}{t}\left[\frac{L'-W'}{W} + \frac{W'}{W-W'}\right] = \frac{\rho}{t}\left[\frac{(L'-W')(W-W')+W'W}{W(W-W')}\right]$$

Assuming $L = iL'$, and normalizing to the original interconnect

$$\frac{R'}{R_0} = \frac{i\frac{\rho}{t}\left[\frac{(L'-W')(W-W')+W'W}{W(W-W')}\right]}{\frac{\rho L}{tW}}$$

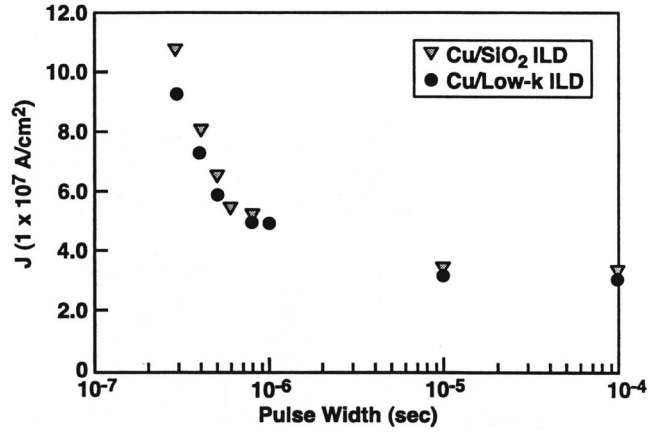

Figure 9.26 Critical current density to failure for Cu interconnect in SiO_2 and low-k ILD

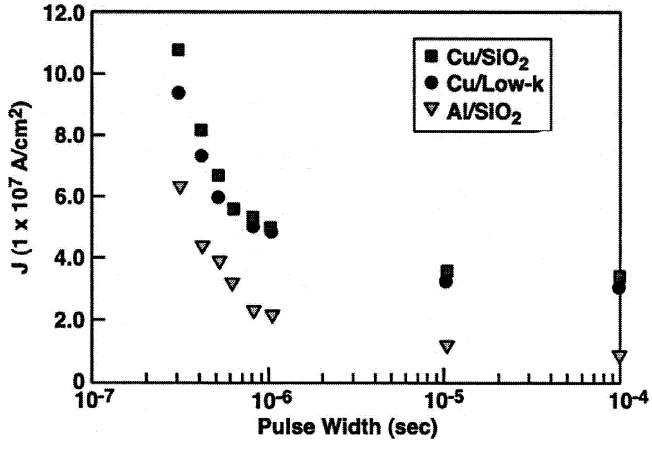

Figure 9.27 Comparison of critical current density to failure of Al and Cu interconnects in SiO_2 and low-k ILD

This can be put in the form

$$\frac{R'}{R_0} = 1 - \left(i \frac{W}{L} \left[\frac{W'^2}{W(W - W')} \right] \right)$$

In this form, the resistance is a function of the number of holes placed in the interconnect as a function of length where i is the parameter of the periodic frequency in the line length for a given width. This model can be expanded to a plurality of shapes in the metal line in the width direction. This can be achieved by modifying the expression,

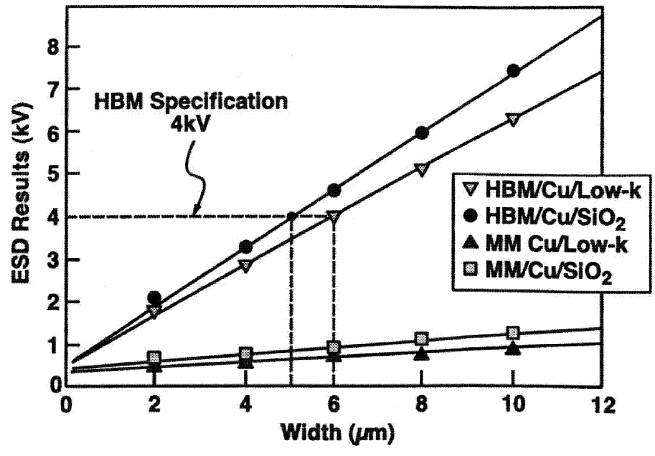

Figure 9.28 ESD linewidth design curves for Cu interconnects in SiO_2 and low-k materials

letting the total W' be jW' where j is the parameter of the periodicity in the width parameter. The new width is then easily integrated into the physical models as

$$R(T) = R'(1 + \alpha T)$$

where $R(T)$ is the dynamic resistance at the temperature T, R' is the initial resistance with the holes in the line, and α is the temperature coefficient of resistance (TCR). From this form, we can solve for temperature in the interconnect, as

$$T = \frac{R(T) - R'}{\alpha R'} = \frac{1}{\alpha}\left(\frac{\Delta R}{R'}\right)$$

Since the impedance is constant, and integration of the expression from the initial to final resistance

$$\frac{\Delta R}{R} = \frac{R_f - R'}{R'} = \alpha\theta_{TH}P$$

Hence, the normalized resistance change is proportional to the product of the thermal impedance, power and temperature coefficient of resistance. Combining Joule heating ($P = I^2R$) and resistance ($R = \rho L/A$), and substituting in for the current density J, then the expression of normalized differential resistance can be expressed as

$$\frac{\Delta R}{R} = \alpha\theta_{TH}J^2\rho LA$$

From the normalized differential resistance, we can solve for the thermal impedance as,

$$\theta_{TH} = \frac{1}{J^2}\frac{[\Delta R/R]}{\alpha\rho LA}$$

In a damascene copper interconnect, the interconnect cross-section is rectangular. The liner film exists on three sidewalls, but on the top region. Since the interconnect is a composite film, with the refractory metal being of a high resistance value, an effective resistivity model can be used where we assume the effective resistivity is

$$\rho_{\text{eff}} = \frac{\rho_{\text{Cu}}}{(1 - 2\Delta/x)(1 - 2\delta/y)} = \frac{\rho_{\text{Cu}}}{\aleph}$$

where the geometrical dimensions are the liner sidewall thickness Δ, liner bottom, δ, and linewidth x, and line height y. We can define an expression which is the normalized change in resistance as

$$\gamma_{\text{crit}} = \frac{(R_{\text{f}} - R')}{R'}$$

and express the critical energy according to the normalized change in resistance at failure,

$$E_{\text{crit}} = \frac{1}{2}[R'(2 + \gamma_{\text{crit}})]I^2\Delta t$$

Energy can be related to the heat capacity and heat of fusion

$$E = C\Delta T + m_i L_{\text{f}}$$

Hence, when the temperature is the critical temperature of failure, then the energy expression is the critical energy. From the resistance relationship on temperature, we can substitute in for the temperature change as a function of normalized resistance change, and the temperature coefficient of resistance

$$E_{\text{crit}} = C\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}$$

In the case of the composite film, the mass of each region, and specific heat of each film must be defined. From the two forms of the critical energy

$$I_{\text{crit}}{}^2 = \frac{\sum_i m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}}{\frac{1}{2}\Delta t R'(2 + \gamma_{\text{crit}})}$$

Expanding this expression to address the multiple films (e.g. copper, cladding and insulator materials) and adding in the hole correction factor

$$I_{\text{crit}}{}^2 = \frac{\left[(m_i c_i + m_j c_j)\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}} + m_k c_k \left(\frac{\gamma_{\text{crit}}}{\alpha}\right)\right]}{\frac{1}{2}\Delta t\left(\rho_{\text{eff}} \frac{L}{A}\left[1 - i\frac{W}{L}\left\{\frac{W'^2}{W(W - W')}\right\}\right]\right)(2 + \gamma_{\text{crit}})}$$

In this expression, the cross-sectional area of the interconnect can be brought into the numerator in order to express the equation as a function of the linewidth and the

geometric variables associated with the films. The physical volume of the Cu region can be expressed as

$$V_{\mathrm{Cu}} = LWd_{\mathrm{s}}\aleph$$

and

$$\aleph = \left(1 - \frac{2\Delta}{W}\right)\left(1 - \frac{2\delta}{d_{\mathrm{s}}}\right)$$

for no hole in the interconnect. In the case of the hole in the interconnect, the copper volume is reduced by the volume of the hole as welll as a correction for the liner on the edge of the hole region.

The cladding volume can be expressed as

$$V_{\mathrm{Clad}} = L\left[W\left(1 - \frac{2\Delta}{W}\right)\delta + 2d_{\mathrm{s}}\Delta\right]$$

With the presence of the hole in the interconnect, the cladding volume must be adjusted for the hole.

The fully passivated thermal sheath volume can be expressed as

$$V_{\mathrm{sheath}} = 4[W + d_{\mathrm{s}} + L]{d_{\mathrm{ox}}}^2 + 2[LW + Ld_{\mathrm{s}} + 2Wd_{\mathrm{s}}]d_{\mathrm{ox}}$$

and the time-dependent oxide sheath thickness is a function of the pulse width, and the heat diffusion coefficient in the oxide

$$d_{\mathrm{ox}}(t) = a_{\mathrm{d}}\sqrt{\Delta t}$$

In this formulation, the sheath volume must include the number of holes, and the volume of the holes. Unlike the volume surrounding the interconnect, the heat is diffusing into the volume from the four sidewalls, significantly reducing the volume absorbing the heat capacity. This volume will not increase with the thermal sheath volume when the sheath extends greater than $W'/2$ (e.g. it is a constant when $d_{\mathrm{ox}} > W'/2$).

9.5 FILL SHAPES AND INTERCONNECTS

Planarization of insulator and conductive films is important in environments with multiple metallization levels. Planarity is achieved using chemical mechanical polishing. To provide good planarity, metal shapes are placed in wide areas. These 'fill shapes' are metal conductors using the metallization level materials. These fill shapes can influence the ESD robustness of interconnects for high-density interconnects and for long-time-constant pulses. The fill shapes can be addressed in the interconnect models by inclusion in the thermal sheath model used for evaluation of copper interconnects. The fill shapes also influence the thermal impedance relative to the bulk substrate. From the development

on Cu interconnects in the case of the composite film, the mass of each region, and specific heat of each film must be defined over all regions, the critical energy to failure is

$$E_{\text{crit}} = \sum m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}$$

From the two forms of the critical energy

$$I_{\text{crit}}{}^2 = \frac{\sum\limits_i m_i c_i \left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}}}{\frac{1}{2}\Delta t R_{\text{o}}(2 + \gamma_{\text{crit}})}$$

Expanding this expression to address the multiple films (e.g. copper, cladding and insulator materials)

$$I_{\text{crit}}{}^2 = \frac{\left[\left(m_i c_i + m_j c_j\right)\left(\frac{\gamma_{\text{crit}}}{\alpha}\right) + m_i L_{\text{f}} + m_k c_k \left(\frac{\gamma_{\text{crit}}}{\alpha}\right)\right]}{\frac{1}{2}\Delta t\left(\rho_{\text{eff}} \frac{L}{A}\right)(2 + \gamma_{\text{crit}})}$$

In this expression, the cross-sectional area of the interconnect can be brought into the numerator in order to express the equation as a function of the linewidth and the geometric variables associated with the films.

The physical volume of the Cu region where conduction is taking place can be expressed as

$$V_{\text{Cu}} = LWd_{\text{s}}\aleph$$

and

$$\aleph = \left(1 - \frac{2\Delta}{W}\right)\left(1 - \frac{2\delta}{d_{\text{s}}}\right)$$

The cladding volume can be expressed as

$$V_{\text{Clad}} = L\left[W\left(1 - \frac{2\Delta}{W}\right)\delta + 2d_{\text{s}}\Delta\right]$$

As the thermal sheath expands, the fill shapes must be included in the thermal stored energy. Each of these fill shapes is of the same physical height, but will have different length and width parameters. The volume of copper is the length of the conducting wire, and the number of copper fill shapes at different levels in the thermal sheath. For the same metal level, the thickness is identical where the number of shapes included in the thermal sheath. This is calculated based on the spacing of the shapes (e.g. the line and space dependence). As the thermal sheath progresses through the insulator volume, the number of adjacent fill shapes are to be included in the width. In the vertical direction,

the number of shapes included is dependent on the thermal sheath and the vertical spacing between successive films and the height of the film.

$$V_{\mathrm{Cu}} = LWd_{\mathrm{s}}\aleph + \sum_{m=1}^{m\leqslant d_{\mathrm{ox}}/P_{\mathrm{m,m-1}}} \mathrm{d}_{sm}\aleph_m \sum_{n=1}^{n\leqslant d_{\mathrm{ox}}/P_{\mathrm{n}}} L_{mn}W_{mn}$$

where the summation variable 'm' and 'n' are associated with the interconnect levels, and 'P' is the pitch between successive levels and lateral pitch between the fill shapes contained in the sheath.

The cladding can also be expressed similarly, including is the cladding of the current-carrying interconnect and that of the number of fill shapes within the thermal sheath.

The fully passivated thermal sheath volume can be expressed as

$$V_{\mathrm{sheath}} = 4[W + d_{\mathrm{s}} + L]{d_{\mathrm{ox}}}^2 + 2[LW + Ld_{\mathrm{s}} + 2Wd_{\mathrm{s}}]d_{\mathrm{ox}}$$

without the fill shapes included. Including the fill shapes the number of fill shapes can be extracted from the volume of the physical sheath

$$V_{\mathrm{sheath}} = (L + 2d_{\mathrm{ox}})(W + 2d_{\mathrm{ox}})(d_{\mathrm{s}} + 2d_{\mathrm{ox}}) - LWd_{\mathrm{s}}\aleph - \sum_{m=1}^{m\leqslant d_{\mathrm{ox}}/P_{\mathrm{m,m-1}}} \mathrm{d}_{sm} \sum_{n=1}^{n\leqslant d_{\mathrm{ox}}/P_{\mathrm{n}}} L_{mn}W_{mn}$$

and the time-dependent oxide sheath thickness is a function of the pulse width, and the heat diffusion coefficient in the oxide

$$d_{\mathrm{ox}}(t) = a_{\mathrm{d}}\sqrt{\Delta t}$$

Let the region be such that the indices i, j, k be in the order of Cu, cladding and the insulator volume, respectively.

Expressing this term as a function of the above volumes, heat capacity and density

$${I_{\mathrm{crit}}}^2 = \frac{2Wd_{\mathrm{s}}\left[\left(\psi_i V_i C_i + \psi_j V_j C_j\right)\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right) + \psi_i V_i L_{\mathrm{f}} + \psi_k V_k c_k\left(\frac{\gamma_{\mathrm{crit}}}{\alpha}\right)\right]}{\Delta t(\rho_{\mathrm{eff}}L)(2 + \gamma_{\mathrm{crit}})}$$

The above model addresses the change in the thermal heat capacity of the insulator sheath, but does not address the correction to the time response of the media with the presence of the fill shapes.

9.6 COPPER FILM STRESS AND ELECTROMIGRATION

ESD stressing of interconnects can lead to material changes and stress in the copper interconnects. Electrostatic discharge events can degrade the electromigration (EM) reliability of interconnect systems as was observed in aluminum interconnects [40]. In Cu interconnects, the EM lifetime has a different character compared with Al-based interconnects [41]. In Cu interconnects, Khoo *et al.* [42] demonstrated the change in the EM lifetime as well as noted the change in the microstructure as a result of TLP testing. They showed that the material property of the Cu film changed after TLP stressing.

From the TEM imaging of stressed and unstressed Cu lines, the stressed Cu stripe shows reduction in the average Cu grain size compared with the unstressed Cu stripe.

With the Cu grain size expected to be less than 0.2 μm, typical 'bamboo' structure is observed in the unstressed line; this shows a marked difference from the numerous small grains observed in the stressed line. When the Cu interconnect is subjected to a high-current pulse, heating occurs that increases the effective volume of the Cu film. However, no significant resistance shift is observed before and after TLP stress. A short pulse duration may not allow for any possibility of void formation to reach a size which could be detrimental to the initial resistance. It is also possible that a reduced average grain size of the Cu grains due to the high-current stress has resulted in early EM failures as a function of the TLP current pulse. This is further confirmed with the diffraction patterns obtained when comparing both a stressed and unstressed interconnect (Figure 9.29). As the Cu grains become smaller due to the heating and subsequent recrystallization, any diffraction rings in the TEM analysis are expected to show more diffusion with increasing levels of applied ESD stress. Figure 9.29 shows the diffraction pattern of an unstressed Cu stripe [43].

For fine-grained polycrystalline structures, diffraction patterns obtained from TEM will appear as a series of concentric rings. As each crystal produces a few diffracted beams, the coexistence of many crystals will generate continuous lines. When grain structures get larger, they exhibit spot patterns. The spots correspond to magnified images of planar sections through the reciprocal lattice, taken normal to the incident beam direction. Any illuminated area shows a single crystal with diffraction spots for several well-defined low-index beam directions. The spot pattern obtained in Figure 9.30 is evidence that larger Cu grains exist in an unstressed interconnect. Figure 9.30 is the diffraction pattern of a TLP-stressed Cu stripe. Comparatively, the ring patterns in the stressed interconnect reflect the presence of smaller-grain polycrystalline material.

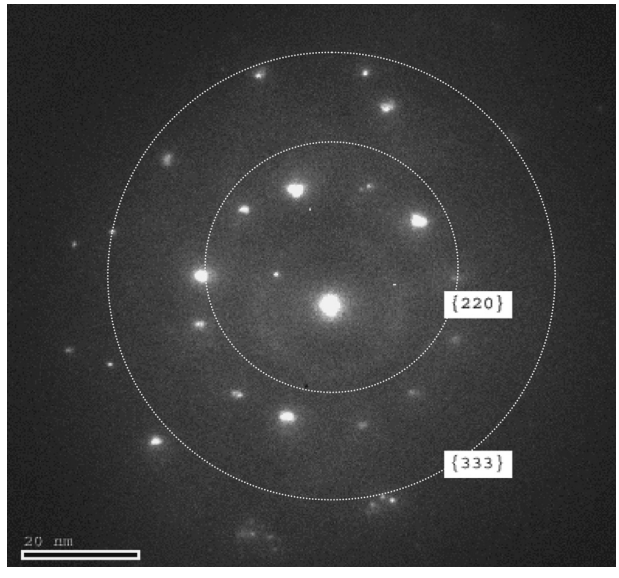

Figure 9.29 Diffraction pattern of an unstressed Cu stripe

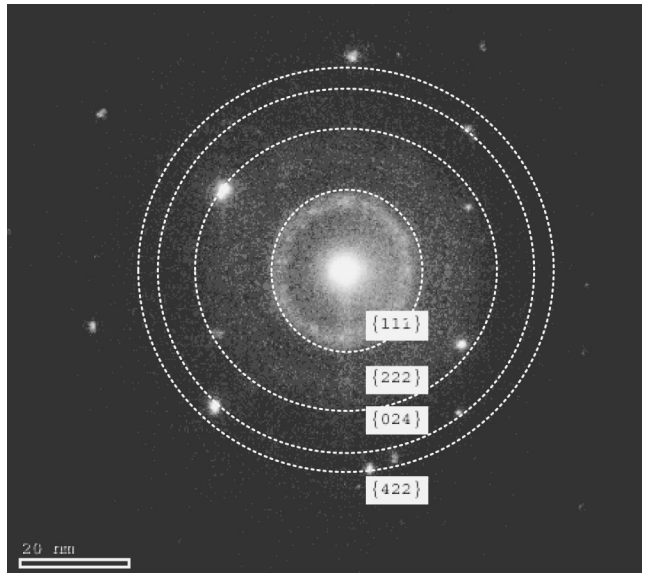

Figure 9.30 Diffraction pattern of TLP-stressed Cu stripe

In accordance with Bragg's law

$$\lambda = 2\ d_{\mathrm{hkl}} \sin(\theta_{\mathrm{B}})$$

Since λ is very small and for a TEM, λ is constant, it is possible to determine d_{hkl} by measuring $\theta_{\mathrm{B's}}$ (the Bragg angle).

Black's model for EM lifetime prediction is stated as

$$T_{50} = AJ^{-n} \exp\left\{\frac{E_{\mathrm{a}}}{k\mathrm{T}}\right\}$$

where J is the current density, E_{a} is the activation energy, and kT is the product of the Boltzmann constant and temperature in Kelvin. Khoo *et al.* [43] showed that, as the prestress TLP current increased, the T_{50} EM lifetime decreased after electromigration stress [43].

During ESD stressing of the Cu interconnect, it was shown that resistance shifts can occur in the Cu film. From the work of Voldman, material changes were also observed [28–30]. From the results of Khoo *et al.*, the stress of interconnects from ESD events have been shown to lead to a latency effect which may manifest itself as EM failures.

A reciprocal relationship between d_{hkl} and θ_{B} is obtained. The element Cu is a known face-centered cubic (FCC) material, it was thus possible to index the diffraction patterns obtained by calibrating it to a Si $<110>$ pattern. In this case, the spots and rings could be indexed, but no obvious new phases could be identified.

The results are similar to previously published TEM microanalysis by Banerjee of TLP-stressed aluminum stripe films [39]. As in the prior work, a smaller grain size is evident in the TLP-stressed Cu interconnect line. Hence, it is concluded that the TLP-stress-induced smaller grain size accounts for the lower EM lifetime.

However, in our experimental results, it is shown that the Cu line EM lifetime degradation of the worst case (F8) is only about 1.4 times that of the unstressed device (F0). Thus EM lifetime of Cu metal is better as compared with Al metal reported in previous work [2] with AlCu EM lifetime being reduced by a factor of 4. In the previously published work of Voldman, it is shown that the Cu line/SiO_2 ILD system is more resilient to ESD stress from human body model (HBM), machine model (MM) and TLP stress compared with the Al line/SiO_2 ILD system. Additionally, Voldman showed that the Cu interconnects/low-k ILD film system has a lower ESD robustness compared with the Cu line/SiO_2 ILD system, but superior to a to the Al line/SiO_2 ILD system. The results are consistent with the previously published work of Voldman in that they show that Cu interconnects/FSG ILD systems are more resilient to ESD compared with Al interconnects/SiO_2 ILD systems.

9.7 INTERCONNECT FAILURE AND VOIDING

During electrical pulsing of interconnects, the thermal melting of the interconnect can lead to material voiding and change of the material structure. In the region of the void, the material is altered, leading to resistance shifts. From the pulsing of copper interconnects, the onset of voiding and material change occurs in a periodic fashion.

Given an interconnect of initial resistance R_0, width W, length L, film thickness t, and resistivity ρ, we can define a void is initiated in the interconnect of width W' and length L'. The size of the void is narrower than the width of the interconnect. Hence, we can assume that $W > W'$. Let us assume that the region of the interconnect failure is a region of a damaged resistivity [44].

Assuming a periodicity of voids k, we can express the damaged line as

$$R' = k\frac{\rho \mathrm{d}}{t}\left[\frac{W}{W-W'}\right] + k\frac{\rho_{\mathrm{u}}}{t}\left[\frac{L'-W}{W}\right]$$

where the first term is the number of damaged regions of length W, the effective width is $W'-W'$, film thickness t, and the damaged sheet resistivity is ρ_{d}.

The second term is the length of the regions of the undamaged sections. Normalizing this expression to the undamaged wire of length L

$$\frac{R'}{R} = \frac{k\dfrac{\rho_{\mathrm{d}}}{t}\left[\dfrac{W}{W-W'}\right] + k\dfrac{\rho_{\mathrm{u}}}{t}\left[\dfrac{L'-W}{W}\right]}{\rho_{\mathrm{u}}\dfrac{L}{tW}}$$

Hence

$$\frac{R'}{R} = k\frac{W}{L}\left(\frac{\rho_{\mathrm{d}}}{\rho_{\mathrm{u}}}\right)\left(\frac{W}{W-W'}\right) + k\frac{W}{L}\left(\frac{L'}{W}\right) - k\frac{W}{L}$$

and substituting $L' = L/k$, we obtain the form

$$\frac{R'}{R} = k\frac{W}{L}\left[\left(\frac{\rho_{\mathrm{d}}}{\rho_{\mathrm{u}}}\right)\left(\frac{W}{W - W'}\right) - 1\right] + 1$$

In the preceeding sections, the change in resistance was during the pulse event, whereas this expression is associated with a permanent change in the resistance after the pulsing. Note that the dynamic resistance (where R' is the d.c. initial resistance) can be evaluated as

$$T = \frac{R(T) - R'}{\alpha R'} = \frac{1}{\alpha}\left(\frac{\Delta R}{R'}\right)$$

Since the impedance is constant, by integration of the expression from the initial to final resistance

$$\frac{\Delta R}{R} = \frac{Rf - R'}{R'} = \alpha\theta_{\mathrm{TH}}{}^{P}$$

In a damaged line, a latent mechanism of concern is the resistance-noise power spectrum. The $1/f$ noise in a metal line is found from Hooge's law [44], which can be expressed as

$$S_R = \frac{1}{I^4 f}\int \{C_{\mathrm{us}}\rho^2 J^4\}\mathrm{d}V$$

where S_R is the resistance-noise power spectrum, I is the current, f is the frequency and C_{us} is the characteristic noise per unit volume. In the ESD damaged interconnect, this can be expressed

$$\frac{S_{R'}}{S_R} = 1 + k\frac{W}{L}\left[\left(\frac{\rho_{\mathrm{d}}}{\rho_{\mathrm{u}}}\right)^2\left(\frac{W}{W - W'}\right)^3 - 1\right]$$

9.8 INSULATOR MECHANICAL STRESS

In ESD events, dielectrics and metal films can undergo mechanical failure due to the thermal stresses induced by strong thermal gradients. Compressive, tensile and shear stresses can lead to mechanical failure in the dielectrics, metal stripes, and via structures.

In the elastic regime, stress τ is proportional to strain ε, where the proportionality constant is Young's modulus, E.

$$\tau = E\varepsilon$$

The point at which this relationship is no longer valid is termed the ultimate stress. During ESD events, Joule self-heating occurs in conductors, and semiconductors as

a result of the resistance in the structures. The Joule heating that occurs in the conductors, and semiconductors leads to thermal strain,

$$\varepsilon = \alpha \Delta T$$

The stress induced by self-heating in the elastic regime is

$$\tau = E\alpha \Delta T$$

From thermal physics, the change in temperature due to self-heating can be expressed as the product of the thermal impedance, and the input power in the structure

$$\Delta T = \Theta_{\mathrm{TH}} P$$

Then, simplistically, the self-heating induced thermal stress can be expressed as

$$\tau = E\alpha \Theta_{\mathrm{TH}} P$$

Mechanical failure of interconnect structures occurs at the ultimate yield conditions. In materials, the yield condition at a point is independent of hydrostatic stress at that point. Hydrostatic mechanical stress does not involve shear stress. Materials that are Hookean in nature do not exhibit shear strain.

The Tresca criterion for the onset of mechanical yielding occurs at a stress condition whose magnitude exceeds one half of the difference between the maximum and minimum principal stress

$$\tau_{\mathrm{Tresca}} = \frac{1}{2}(\tau_{\max} - \tau_{\min})$$

The Mises–Hencky criterion defines a shear stress known as the octahedral shear stress where the octahedral shear stress is expressed as a function of the principal stresses [45]

$$\tau_{\mathrm{OCT}} = \frac{1}{3}\sqrt{(\tau_1 - \tau_2)^2 + (\tau_1 - \tau_3)^2 + (\tau_2 - \tau_3)^2}$$

When the stress in a material at a point exceeds the octahedral shear stress, material yielding can occur. The Mises–Hencky criterion has been shown to be valid for ductile materials such as copper and aluminum while the Tresca criterion is more appropriate to brittle materials. In semiconductor components, to address the yielding, the prestress in the films must be considered with the additional thermal stress to assess the mechanical yielding during an ESD event.

During ESD events, mechanical stress occurs in the interconnect dielectric films due to the high thermal stresses induced by the ESD pulse. The evolution of the failure mechanism of the interconnect–dielectric system is related to the dielectric material, the interconnect cladding material and location, and the interconnect film material. The location of the cladding material is important in the evolution of the interconnect–dielectric system. For example in aluminum interconnects, the titanium cladding is

placed on top and bottom of the aluminum film. In the evaluation of the interconnect failures, a first observation is the cracking of the insulator on the edges of the interconnect. After interconnect cracking, the aluminum flows into the laterally cracked areas. From a top observation, the original cladding material will remain intact due to the higher melting temperature. Cladding failure is typically observed after the observation of the aluminum film in the dielectric area. After successive ESD pulses, if electrical connectivity is maintained, the current will flow through the cladding material, leading to elephantine patterns formed in the cladding film, followed by a separation of the electrical continuity and failure of the cladding material. In a copper damascene process, the cladding is placed on the three edges of the trough region. In this case, the first observation is the change in the microstructure of the Cu film. After the continued ESD stressing, the Cu film melting and insulator cracking is observed. The cracking of the insulator is an upward cracking; the cracking appears to occur at the Cu–cladding interface triple-point, and progresses upward. The insulator interface where the cracking is the surface where no tantulum (Ta) cladding is present. Continued stressing of the interconnect leads to Cu and Ta cladding failure and penetration into the insulator volume.

In Chapter 10, we discuss silicon-on-insulator (SOI) technology and ESD. With the transition to SOI technology, there is some synergy of the role of the interconnects in the heat dissipation. Hence the issues addressed in this chapter, are relevant to bulk or SOI technology. In fact, there are some new interesting issues on the role of the interconnect in this environment.

PROBLEMS

9.1. Derive the ratio of the critical current of an aluminum interconnect to the critical current of a copper interconnect.

9.2. From the above relationship, assume the Cu interconnect is scaled by the MOSFET constant electric field scaling parameter in the film thickness only. Derive the relationship where the Cu interconnect is also scaled in the metal width. Assume the cladding thickness does not scale.

9.3. Derive the ratio of the critical current to failure of a Cu interconnect in SiO_2 compared with a Cu interconnect in a low-k material.

9.4. Assume for a fixed interconnect that the low-k material thermal conductivity is reduced according to the MOSFET constant electric field scaling parameter for each technology generation, derive a relationship showing the critical current to failure assuming a fixed line thickness and width. Derive a relationship where the film thickness and the width of the metal is scaled with the low-k material.

9.5. Assuming 'constant ESD scaling theory' where the ESD robustness of the interconnect system is a constant with technology scaling, derive the metal width relationship of each metal level assuming the metal film thickness is scaled as the MOSFET constant electric field scaling parameter α, where each metal level is thicker by the scaling parameter from the device surface to the last level metal (e.g. the wire interconnects get thicker by α according to the wiring hierarchy concept).

9.6. Derive a relationship for the density of fill shapes which are necessary to compensate the thermal conductivity of the low-k material assuming the fill shapes are Cu material so that the ESD robustness does not decrease with each technology generation.

REFERENCES

1. D. C. Wunsch and R. R. Bell. Determination of threshold failure levels of semiconductor diodes and transistors due to pulsed voltages. *IEEE Transactions on Nuclear Science* 1968; **NS-15** (6): 244–259.
2. D. Tasca. Pulse power modes in semiconductors. *IEEE Transactions on Nuclear Science* 1968; **NS-15** (6): 244–259.
3. D. Tasca. MINUTEMAN III re-entry system piece parts support test (submicrosecond pulse electric failure modes in semiconductor devices). *General Electric Co., Document No. 70SD401*, January 1970.
4. J. Smith. Pulse power testing of microcircuits. *RADC-TR-71-59*, Rome Air Development Center, October 1971.
5. J. Smith. Electrical overstress failure analysis in microcircuits. *Proceedings of the EOS/ESD Symposium*, 1981; 41–46.
6. J. Smith and W. Littau. Prediction of thin film resistor burnout. *Proceedings of the EOS/ESD Symposium*, 1981; 192–197.
7. D. Egelkrout. Metalization failure and thermomechanical shock studies. *Boeing Corporation, Defense Atomic Support Agency Report 2611 (SRD)*, November 1971.
8. E. Kinsbron *et al*. Failure of small thin film conductors due to high current density pulses. *Proceedings of the International Reliability Physics Symposium*, 1978; 248–254.
9. D. Pierce. Modeling metalization burnout of integrated circuits. *Proceedings of the EOS/ESD Symposium*, 1982; 56–61.
10. S. Voldman. The impact of technology evolution and scaling on electrostatic discharge (ESD) protection on high-pin-count high-performance microprocessors. *Proceedings, of the International Solid-State Circuits Conference* Session WA21, San Francisco, 15–17 February 1999; 366–368.
11. S. Voldman. ESD robustness and scaling implications of aluminum and copper interconnects in advanced semiconductor technology. *Proceedings of the EOS/ESD Symposium*, 1997; 316–329.
12. T. Maloney. Integrated circuit metal in the charged device model: bootstrap heating, melt damage, and scaling laws. *Proceedings of the EOS/ESD Symposium*, 1992; 129–134.
13. P. Solomon. The need for low resistance interconnects in future high speed systems. *SPIE Proceedings* 1988; **947**: 104.
14. Y. Mii. Performance consideration for scaling of sub-micron on-chip interconnections. *SPIE Proceedings* 1992; **1805**: 332.
15. G. Sai-Halasz. High end processor trends and limits. *Proceedings of the International Conference on Advanced Microelectronic Devices and Processing*, Sendai, Japan, 1994; 753.
16. G. Sai-Halasz. Performance trends in high end processors. *Proceedings of the IEEE* 1995; **83**: 20.
17. D. Edelstein, G. A. Sai-Halasz and Y. J. Mii. VLSI on-chip interconnection performance simulation and measurements. *IBM Journal of Research and Development* 1995; **39** (4): 383–401.
18. M. Bohr. Interconnect scaling – the real limiter to high performance ULSI. *Solid State Technology* 1996.
19. H. B. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI*. Reading, MA: Addison Wesley, 1990.
20. J. G. Ryan, R. M. Geffken, N. R. Poulin and J. R. Paraszczak. The evolution of interconnection technology in IBM. *IBM Journal of Research and Development* 1995; **39** (4): 371–383.
21. K. Banerjee *et al*. Characterization of VLSI circuit interconnect heating and failure under ESD conditions. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1996; 237–245.

22. D. Edelstein *et al*. Full copper wiring in a sub-0.25 μm CMOS ULSI technology. *International Electron Device Meeting (IEDM) Technical Digest*, Session 33.3.1, December 1997; 773–776.
23. D. Edelstein. Performance and integration of copper VLSI/ULSI interconnects. *Proceedings of the American Vacuum Society 41st National Symposium*, 1994; 123.
24. D. Edelstein. Advantages of copper interconnects. *Proceedings of the 12th International VLSI Multilevel Interconnection Conference*, Santa Clara, CA, 1995; 301.
25. B. Luther *et al*. Planar copper polyimide back end of line interconnections for ULSI devices. *Proceedings of the 12th International VLSI Multilevel Interconnection Conference*, Santa Clara, CA, 1995; 15.
26. C. K. Hu *et al*. Copper–polyimide wiring technology for VLSI circuits. *Materials Research Society Symposium Proceedings VLSI* 1990; **V**: 369.
27. C. Kaanta *et al*. Dual damascene: a ULSI wiring technology. *Proceedings of the 8th International IEEE VLSI Multilevel Interconnect Conference*, Santa Clara, CA, 1990; 123.
28. S. Voldman. ESD robustness and scaling implications of aluminum and copper interconnects in advanced semiconductor technology. *Proceedings of the EOS/ESD Symposium* 1997; 316–329.
29. S. Voldman *et al*. High-current transmission line pulse characterization of aluminum and copper interconnects for advanced CMOS semiconductor Technologies. *Proceedings of the International Reliability Physics Symposium*, 1998; 293–301.
30. S. Voldman. Lightning rods for nanostructures. *Scientific American*, October 2002.
31. SIA Semiconductor Industry Association. *The National Technology Roadmap for Semiconductors*, 1997.
32. SEMATECH Report. *Low Dielectric Constant Materials Characterization and Materials Property Database*, September 1997.
33. L. Peters. Pursuing the perfect low-k dielectric. *Semiconductor International*, September 1998; 64–74.
34. R. K. Laxman Low-ε dielectrics: CVD fluorinated silicon dioxides. *Semiconductor International*, May 1995; 71–74.
35. D. Ambrust *et al*. Integration of HDPCVD FSG for 0.25 μm CMOS Technology. *DUMIC Proceedings* 1998; 67.
36. H. Meynen *et al*. The Integration of low-k hydrogen silsequioxane in sub-0.35 μm processes. *Proceedings of the Electrochemical Society Symposium on Low and High Dielectric Constant Materials*, May 1998.
37. G. Passemard *et al*. Single damascene integration of BCB with copper. *Proceedings of the VLSI Multilevel Interconnect Conference* June 1998; 63.
38. K. Banerjee *et al*. The effect of interconnect scaling and low-k dielectric on the thermal characteristics of the IC metal. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, December 1996.
39. S. Voldman *et al*. High current characterization of dual damascene Copper/SiO_2 and low-K interlevel dielectrics for advanced CMOS semiconductor technologies. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1999; 144–153.
40. K. Banerjee *et al*. Microanalysis of VLSI interconnect failure modes under short-pulse stress conditions. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2000; 283–288.
41. C. Ryu *et al*. Electromigration of submicron damascene copper interconnects. IEEE Symposium of VLSI Technology Digest of Technical Papers, 1998; 156–157.
42. S. C. K. Khoo, P. Y. Tan, E-C. Chua, S-C. T. Carol, I. Manna, S. Redkar and S. Ansari. Copper interconnect microanalysis and electromigration reliability performance due to the impact of TLP ESD. *Proceedings of the EOS/ESD Symposium*, 2002; 382–387.
43. S. Khoo, P. Y. Tan and S. Voldman. Microanalysis and electromigration reliability performance of transmission line pulse (TLP) stressed copper interconnects. *Journal of Microelectronics and Reliability*, 2003; **43**: 1039–1045.

44. L. Chu, W. K. Chim, K. L. Pey and A. See. Effect of transmission line pulsing of interconnects investigated using combined low-frequency noise and resistance measurements. *Proceedings of the International Physical and Failure Analysis Symposium*, 2001; 97–102.
45. I. Shames. *Introduction to Solid Mechanics*. Engdwood cliffs: Prentice Hall, 1975.

10 Silicon On Insulator (SOI) and ESD

This chapter addresses silicon-on-insulator (SOI) technology and ESD. We discuss lateral polysilicon gated diode structures in SOI. SOI diode models, and thermal models are introduced. From the generalist approach, SOI modeling is a natural for Green's function thermal models, since the geometries and the substrate can be treated as a stratified medium problem. The dynamic threshold SOI devices is also introduced to show new opportunities in SOI environments.

Silicon-on-insulator (SOI) was viewed as the natural extension of CMOS technology in order to maintain the performance objectives set up by Moore's law [1]. In the early 1990s, there was a greater interest in SOI as a natural extension of bulk CMOS technology in order to maintain Moore's scaling law. This stemmed from the fact that SOI reduces junction capacitance which lowers the performance objectives. The CMOS gate delay, excluding the intrinsic delay, can be expressed as

$$(T_g)_{\text{extrinsic}} = f_g R_d C_w L_w + 0.4 R_w C_w L_w^2 + 0.7 R_w C_r L_w$$

where the extrinsic gate delay are the interconnect related delay terms R_d and C_d are the MOSFET output resistance and capacitance, respectively, C_r is the MOSFET receiver switching capacitance, R_w, C_w, and L_w are the resistance, capacitance, and the line length of the interconnect, and f_g is the circuit fan-out. As the technology is scaled, the interconnect delays will be reduced by transition to Cu interconnects, and low-k materials. This leaves the fundamental speed of the transistor, associated with the diffusion capacitance and resistance loading effects.

The migration from bulk CMOS to SOI technology was driven by the desire for increased MOSFET performance in advanced microprocessors as MOSFET transistors scaled to channel length dimensions below 0.1 μm. SOI has emerged as a mainstream commercially viable technology for CMOS-based advanced microprocessors [2–7]. For many years, SOI technology was used in applications that needed high immunity to single-event upsets (SEU) for space applications. In space applications,

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

bulk CMOS technology was vulnerable to charge generation from heavy ions, protons, neutrons and other particles. The heavy ions and charged particles produced charge generation tracks as they traversed the silicon lattice. Additionally, charge generation was generated from interaction of the particle and the silicon lattice nucleus. The charge generation led to soft error events in memory cells, logic disturbs in logic circuits and single-event induced latchup events. Bulk CMOS technology was sensitive to both single-event upsets, and latchup in space applications. As a result, SOI remained as a niche market for space and military applications. In recent times, SOI was of interest in radio frequency communications applications. SOI applications provide good noise isolation and low capacitance.

A key concern was whether electrostatic discharge protection would be an impediment to the introduction of SOI as a mainstream technology that would follow standard bulk CMOS technology. Hence, the success of achieving electrostatic discharge protection in SOI technologies was important in the acceptance of SOI as a mainstream technology for commercial semiconductor chips. It was commonly believed that providing ESD protection in SOI technology was a substantial barrier or challenge to overcome in the early 1990s [8–10].

MOSFET scaling of bulk CMOS technology continues to be the trend for high-density high-performance advanced microprocessors. As technologies scale below 0.15 μm technology generation, MOSFET issues, such as short channel effects (SCE) control, gate resistance, channel profiling, and hot electrons, continue to challenge device engineers designing in bulk CMOS technology. Power consumption is a function of the capacitance, power-supply voltage and transition frequency. To reduce power consumption, microprocessor chip power can be lowered by decreasing the power-supply voltage [1]. At lower power-supply voltages, the junction capacitance severely impacts the performance of bulk CMOS devices. Using SOI technology, many MOSFET device concerns and obstacles of bulk CMOS can be eliminated. CMOS-on-SOI has significant advantages over bulk CMOS technology, and will achieve the scaling objectives for high-performance advanced microprocessors. CMOS-on-SOI technology provides low leakage current, low capacitance MOSFETs and resistor elements, and good subthreshold I–V characteristics [2]. But, for SOI to become a manufacturable mainstream technology, the floating body effects and ESD protection must be addressed. Floating body effects are observable from the kink effect, subthreshold currents and MOSFET threshold voltage variations. These effects are solvable by body contacts, ultra-thin SOI and fully depleted MOSFETs. In fully depleted (FD) SOI, the floating body effects are sensitive to SOI film thickness and the depletion region must always be greater than the silicon film thickness. Fully depleted SOI devices do not allow for lateral bipolar conduction, which would be advantageous for ESD protection networks. With partially depleted (PD) SOI technology with body contacts, floating body effects and ESD concerns have been addressed. Using PD-SOI technology, the bipolar current gain is reduced by using silicon films that are not too thick. As the silicon film thickness decreases, the emitter and collector areas are reduced, lowering the bipolar current gain.

In SOI development, new ESD structures are needed to provide ESD protection. Figure 10.1 is an example of a SOI ESD network that will be discussed. In the following sections, SOI models, devices and experimental results will be shown.

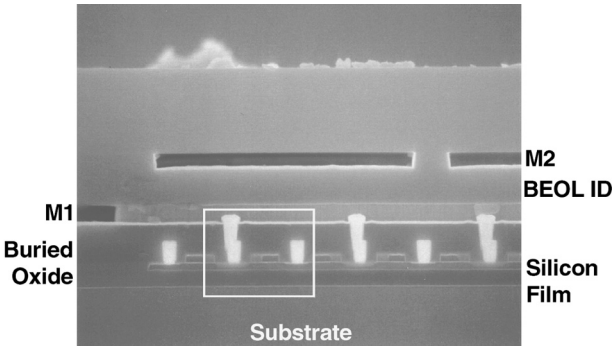

Figure 10.1 Cross-section of an SOI ESD network

10.1 SOI ELECTROTHERMAL MODELS

Electrothermal models for SOI can utilize Green's function techniques because of the geometrical simplicity of SOI structures. In an SOI device, whether a SOI diode, SOI resistor element, or SOI MOSFETs, the physical volume of the conducting region is a parallelepiped with a rectangular geometrical boundaries. The upper surface is an insulator from the dielectric films, the lower surface boundary is the buried oxide (BOX) region, and the edges are typically shallow trench isolation (STI) regions. The solution of the parallelepiped is appropriate for the case of the semi-infinite region, can be modified and the method of images can be applied. The solution due to Joy and Schlig can be utilized [11]. In the simplest solution, we can assume the thermal region is contained within an infinite space of insulators. For short time scales, this is a valid assumption.

A solution where the physical volume V of silicon is rectangular, where the dimensions are length L, width W, and height H, $V = LWH$, and we can define a heat capacity associated with the volume of the parallelpiped, the specific heat of silicon and mass density, where

$$C = \mathrm{c}\rho V$$

In the general case, this parallelepiped can be displaced from the surface. We can show the solution using a SOI parallelepiped of identical dimensions displaced a distance D above the $z = 0$ plane, a plane of symmetry is established that has a solution where all points along the plane $z = 0$ have the heat flux at the surface equal to zero. To solve the problem of the semi-infinite domain, we can utilize the method of images by modification of the Green's function in the infinite domain so that it applies to the semi-infinite domain. From the time-dependent heat equation, with a constant thermal conductivity and constant specific heat–mass density product, the Laplacian of temperature is proportional to the partial derivative of temperature as a function of time. As in the problem of an infinite domain, we have

$$T(x, x'; y, y'; z, z'; t) = \int_{t'=0}^{t'=t} \mathrm{d}t' \frac{Q(t')}{8[\pi\kappa(t - t')]^{3/2}} \exp\left\{-\frac{r^2}{4\kappa(t - t')}\right\}$$

Expressing this as a function of power dissipated in an infinitesimal volume formed between x' and $x'+\mathrm{d}x'$, y' and $y+\mathrm{d}y'$, and z' and $z+\mathrm{d}z'$, and defining the power dissipated as normalized to the volume V

$$T(x,x';y,y';z,z';t)=\frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t}\mathrm{d}t'\frac{\{P(t')\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'\}}{[(t-t')]^{3/2}}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}$$

To evaluate the temperature over all space from all infinitesimal volumes, the expression can be integrated over the infinite volume Ω. In the expression, the spatially independent terms can be removed from the integration over space, and the temperature can be expressed as

$$T(x,y,z;t)=\frac{1}{8\rho cV(\pi\kappa)^{3/2}}\int_{t'=0}^{t'=t}\mathrm{d}t'\frac{P(t')}{[(t-t')]^{3/2}}\int_{\Omega}\exp\left\{-\frac{r^2}{4\kappa(t-t')}\right\}\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$$

For evaluation of a parallelepiped in an infinite medium let us assume a source has the dimension W in the x-direction, L in the y-direction, and H in the z-direction, but applying a parallelepiped displaced distance D below the boundary condition $z=0$ and a image source of equal and opposite strength above the $z=0$ plane at $z=D$

$$T(x,y,z;t)=\frac{1}{8\rho cV(\kappa)^{3/2}}\int_{t'=0}^{t'=t}{}'\frac{P(t')\mathrm{d}t'}{[(t-t')]^{3/2}}F(x-x',y-y',z-z',t-t')$$

with

$$F(x-x',y-y',z-z',t-t')=F_x(x-x':t-t')F_y(y-y';t-t')F_z(z-z';t-t')$$

where

$$F_x(x-x',t-t')=\int_{-W/2}^{W/2}\exp\left\{-\frac{(x-x')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}x'}{\sqrt{\pi}}$$

$$F_y(y-y',t-t')=\int_{-L/2}^{L/2}\exp\left\{-\frac{(y-y')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}y'}{\sqrt{\pi}}$$

$$F_z(z-z',t-t')=\int_{-(D+H)}^{-D}\exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}z'}{\sqrt{\pi}}+\int_{D}^{D+H}\exp\left\{-\frac{(z-z')^2}{4\kappa(t-t')}\right\}\frac{\mathrm{d}z'}{\sqrt{\pi}}$$

The integral can be expressed as error functions using a transformation of variables. The expression for temperature in an infinite medium can be put in the form

$$T(x,y,z;t)=\frac{1}{8C}\int_{t'=0}^{t'=t}P(t')H(x,y,z;t-t')\mathrm{d}t'$$

where the volume $V=LWH$, and letting $C=c\rho V$

we can write the function containing the spatial dependence as the product of the error functions

$$H(x,y,z;t-t') = H(z;t-t') \prod_{i=x,y} \left[erf\left(\frac{(L_{x_i}/2)+x_i}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{(L_{x_i}/2)-x_i}{\sqrt{4\kappa(t-t')}} \right) \right]$$

and

$$H(z;t-t') = \left[erf\left(\frac{z+D+H}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{-D-z}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{z-D}{\sqrt{4\kappa(t-t')}} \right) + erf\left(\frac{D+H-z}{\sqrt{4\kappa(t-t')}} \right) \right]$$

For ESD events, the region of power generation is contained within the physical volume of the device. Power is associated with the dissipated power by the SOI semiconductor component. The power as a function of time is associated with the power dissipated in the device structure, which can be calculated from the actual current and voltage as a function of time through the device structure. The power through the device can be obtained from the transmission line pulse I–V characteristics. In transmission line pulse testing, the input pulse is a rectangular pulse of a given current level. The input pulse can be represented as a rectangular power pulse of a given pulse length. The input pulse can be represented mathematically as a constant absorbed power term times a Heaviside step function associated with a pulse of length τ_P. The absorbed power can be evaluated based on the transmission and reflection coefficients of the TLP system.

To evaluate the peak temperature, a constant power magnitude associated with the Wunsch–Bell power to failure for a given pulse width, $P(\tau_p)$ can be used. This avoids the integration and the thermal response in the medium and temperature rise can be simplistically evaluated.

In all ESD modeling there is judgement of the physical volume to include in the thermal analysis. In the simple case where we assume the volume of interest is not displaced from the surface, $D=0$. In all the structures of interest, the physical volume can be represented by the complete area contained within the isolation regions. For the case of the SOI resistor element, the volume V would be represented as the dimensions of the complete resistor element. In the case of an SOI diode element, the volume could be represented as the metallurgical junction region, and can include the low-doped regions of the structure (e.g. n^- region of the lateral gated SOI $p^+/n^-/n^+$ diode). In an SOI MOSFET, the volumetric region can be modeled as the region of the channel region where the peak heating is undergone. In bulk CMOS, because of the undefined boundary surfaces, typically a parallelepiped is formed near the drain region associated with the MOSFET width, junction depth, and the depletion width at the drain structure. In the SOI implementation, the insulation boundaries are clearly defined, allowing for better judgement of the boundary conditions used for SOI modeling.

In the analysis, we also assume ideal boundary conditions of insulators. Heat diffusion into the insulator volume can be addressed by using mixed boundary conditions at the silicon–silicon oxide interfaces. A simpler method is to assume a time-dependent

thermal sheath volume of insulator around the silicon region, as was utilized in the interconnect model developments in the Chapter 9.

10.1.1 SOI Electrothermal Transmission Line Models

Electrothermal models for SOI under ESD conditions are highly dependent on the evaluation of the physical volume where the self-heating in the SOI device is occurring, as well as the thermal boundary conditions. Wherein bulk CMOS, lossy transmission line models were proposed to model the substrate region, in SOI technology, the role of the thermal boundary conditions from the vertical stratified medium and the thermal contact boundary conditions will have a strong influence on the prediction of the power to failure of the SOI structure from the ESD event.

From the general form of an electrical lossy transmission line model, we can express the temperature and thermal flux relationship as conditions as

$$T(y) = T_1 \exp\left\{-\frac{y}{\gamma}\right\} + T_2 \exp\left\{-\frac{L-y}{\gamma}\right\}$$

Assuming a transmission line with a termination thermal resistance R_1 at terminal 1 and a termination thermal resistance R_2 at terminal 2, we can express the transmission line system in the form of a matrix analogous to the electrical model proposed by Troutman for the bulk substrate (see Chapter 4, Section 4.5, on lossy transmission lines). From the Troutman formulation, the lossy transmission time model can be applied to the ESD event using thermal terminations instead of the electrical analog. In this form, the temperature can be related to the thermal impedance in a two-port matrix formulation.

Raha, Ramaswamy and Rosenbaum [12] proposed the use of a thermal RC network as a thermal transmission line model for simulation of ESD events in SOI devices. From the parallelepiped SOI model, the thermal generation can serve as a time-dependent power generation source term. In the model, the transmission line equation can be represented as a thermal analog

$$\frac{\partial^2 T(x,t)}{\partial x^2} - RY(t)T(x,t) = Rg(t)$$

The transmission line equation is second order in space with a time-dependent variable coefficient, and a generation drive term associated with the ESD pulse. The admittance $Y(t)$ is associated with the thermal network consisting of the vertical stratified medium.

$$Y(t) = \sum_i Y(t)$$

In this form, the thermal model addresses the relationship of the power and the temperature. The node of interest is the temperature in the semiconductor device. The regions outside of the device regions can serve as thermal resistance and thermal capacitance terminations of the thermal transmission line.

Treating the ESD events as step impulse of thermal power generation at time $t = 0$, the analysis can be evaluated in the frequency domain. This can be transformed into the Laplacian domain where for a step function in the thermal power generation we can express the equation as

$$\frac{\partial^2 T(x,s)}{\partial x^2} - RY(s)T(x,s) = R\frac{g_0}{s}$$

The thermal power generation can be related to the power generated inside the semiconductor structure divided by the volumetric region of the semiconductor element. Raha *et al.* chose the power density as

$$g_0 = \frac{Power\ to\ failure}{V} = \frac{V_{t2}I_{t2}}{WLx_{\mathrm{d}}}$$

where the power density is the power to failure over the volume of the power generation. In a MOSFET structure, the peak power to failure is the product of the MOSFET second breakdown current and MOSFET second breakdown voltage divided by the MOSFET depletion region.

To evaluate the solution of the thermal transmission line model, the boundary conditions at the ends of the thermal transmission line must be evaluated. The thermal termination terms of the transmission line are the metal interconnects and inter-level dielectric films, and the substrate region. These terminations influence the thermal response and peak temperature from the generation term. The doping concentration will influence the thermal admittance of the substrate region and the contacts to the substrate. For the other termination, the salicide, the contact resistance, interconnects and the thermal conductivity of the interlevel dielectric will influence the magnitude of the thermal dissipation.

Raha *et al.*, in application to a SOI MOSFET structure used the thermal admittance term as the series of the thermal resistance–thermal capacitance RC networks of the depletion region, the neutral region of the thin silicon region, the buried oxide and the silicon substrate when applied to a partially depleted MOSFET structure [12]. The assumption in the analysis is the substrate temperature does not vary below the buried oxide film for short time scales, and the thermal admittance of the metallization and interlevel dielectrics are negligible.

10.1.2 SOI Electrothermal Transmission Line Model Series Solution

From the SOI electrothermal transmission line model, the thermal generation can serve as a time-dependent power generation source term. In the model, the transmission line equation can be represented as a thermal analog [12]

$$\frac{\partial^2 T(x,t)}{\partial x^2} - RY(t)T(x,t) = Rg(t)$$

where admittance $Y(t)$ is associated with the thermal network consisting of the vertical stratified medium.

$$Y(t) = \sum_i Y(t)$$

Treating the ESD events as step impulse of thermal power generation at time $t = 0$, the analysis can be evaluated in the frequency domain. This can be transformed into the Laplacian domain where for a step function in the thermal power generation we can express the equation as the following,

$$\frac{\partial^2 T(x,s)}{\partial x^2} - RY(s)T(x,s) = R\frac{g_0}{s}$$

Assuming that the generation term is in the spatial region between x_1 and x_2, we can provide a series solution. The thermal power generation can be related to the power generated inside the semiconductor structure divided by the volumetric region of the semiconductor element. Raha *et al.* chose the power density as

$$g_0 = \frac{power\ to\ failure}{V} = \frac{V_{t2}I_{t2}}{WLx_{\mathrm{d}}}$$

where the power density is the power to failure over the volume of the power generation. In a MOSFET structure, the peak power to failure is the product of the MOSFET second breakdown current and MOSFET second breakdown voltage divided by the MOSFET depletion region.

From Wang, Juliano and Rosenbaum [13], the series solution can be represented as

$$T(x,t) = \sum_{n=1,3,5}^{\infty} \sin\left(\frac{n\pi x}{A}\right)\Gamma_n(t) + \sum_{n=0,2,4}^{\infty} \cos\left(\frac{n\pi x}{A}\right)\Gamma_n(t)$$

where

$$\Gamma_n(s) = \frac{RD_n}{s[\gamma_n^2 + RY(s)]}$$

for all integers $n = 0, 1, 2, 3\ldots$
and

$$\gamma_n = \frac{n\pi}{A}$$

and

$$D_0 = \frac{q_0(x_2 - x_1)}{A}$$

and

$$D_n = \frac{2q_0}{n\pi}\left[\cos\left(\frac{n\pi x_2}{A}\right) - \cos\left(\frac{n\pi x_1}{A}\right)\right]$$

for odd values of n

$$D_n = \frac{2q_0}{n\pi}\left[\sin\left(\frac{n\pi x_1}{A}\right) - \sin\left(\frac{n\pi x_2}{A}\right)\right]$$

for even values of n, and the source generation term.
where

$$q_0 = \frac{Q}{WC(x_2 - x_1)}\left[\frac{C_2}{C_1 + C_2}\right]$$

where Q is the total heat, divided by the volume of heating (W is the width, $x_2 - x_1$ is the region of heat generation, c is depth of the region of heating within the silicon film), and the two series thermal capacitance terms associated with the silicon film outside the depth c, and the buried oxide thermal capacitance.

10.2 SOI ESD DIODES AND DEVICES

10.2.1 Abrupt Junction Technology

The first introduction of mainstream CMOS technology was abrupt junction CMOS technology. SOI ESD protection networks were first constructed in a 2.5 V CMOS logic technology with a 0.25 μm effective channel length L_{eff} technology [14–20]. Source/drain features were defined using shallow trench isolation (STI), which abuts the silicon dioxide (SiO_2) buried oxide film. Device channel lengths were obtained using a 0.5 numerical aperture (NA) tool and negative-tone resist system. Dual work function silicided polysilicon gate electrodes were used for the MOSFET gate conductors. p-Channel transistors use abrupt boron junctions; n-channel devices have abrupt non-LDD arsenic junctions. MOSFET source/drain junction depths for both the n-channel and p-channel MOSFETs are 0.18 μm. A single-species boron p^+ implant process of 40 keV energy, $5 \times 10^{14}/\text{cm}^2$ dose of Ge preamorphization and a 7 keV $4 \times 10^{15}/\text{cm}^2$ dose of ^{11}B source/drain implant is used to form the p^+ diode. A $TiSi_2$ salicide film is formed on the source/drain junctions, as discussed in Chapter 7. The salicided polysilicon gate structure is placed on a 7.7 nm SiO_2 gate dielectric.

In the bulk implementation, diodes were formed using a 0.55 μm shallow trench isolation and a 1.2 μm p^+ to n^+ spacing. These structures were modified to form polysilicon-defined diode ESD devices [21]. The progression from the polysilicon-defined diodes in bulk CMOS allowed easy migration to SOI technology. For the SOI technology, a lateral $p+/n^-/n+$ gated diode structure was introduced with no additional masking steps. In this implementation, the p^+ junction is an abrupt junction and the $n+/n^-$ transition also uses an abrupt junction (Figure 10.2).

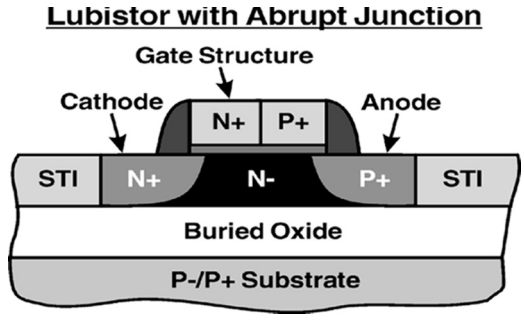

Figure 10.2 Cross-section of an SOI polysilicon gated *p–n* diode (lubistor) ESD structure with abrupt junctions

The mask to define the p^+ and n^+ implants must be placed over the polysilicon gate structure. In this fashion, the polysilicon-bound diode structure has a polysilicon film with two different dopant types and work functions along the device channel length [14–20]. Placement of this structure on a buried oxide, the trench isolation abuts the buried oxide film, isolating the polysilicon diode structure from adjacent structures. In the 0.25 μm technology, the lateral polysilicon diode structure was studied as a function of the SOI diode perimeter and the polysilicon channel length.

In this case, the ESD structure consists of a p^+ diode junction in series with the $n^-/n+$ cathode region. The gate electrode is not activated or used to modulate the current but serves as a 'mask' to allow shorting of the anode and cathode region with salicide. This structure can be constructed as a $p^+/p^-/n^+$ SOI diode structure using the *p*-well implant instead of the *n*-well implant. On the SOI diode anode in an SOI $p^+/n^-/n^+$ diode, the resistance can be expressed as

$$R_{\text{anode}} = R_{\text{M}} + R_{\text{C}} + (R_{\text{Sal}})_{\text{eff}} + (R_{p+})_{\text{eff}}$$

where the anode resistance is the resistance of the interconnect, the contact, the salicide film, and the anode implant. For the silicide film, the material, area coverage and design of the silicide formed on the junction can alter the effectiveness of the silicide as well as the diffusion area on the ESD protection. For SOI structures, because of the presence of the buried oxide film, the role of the metal, contacts and salicide as a source for thermal dissipation is more critical.

Equivalently, for the cathode, the cathode resistance is the sum of the resistances in series as well as the *n*-body resistance. For the cathode, typically the *n*-body resistance is the largest component of resistance which influences the ESD results.

$$R_{\text{cathode}} = R_{\text{M}} + R_{\text{C}} + (R_{\text{Sal}})_{\text{eff}} + (R_{n+})_{\text{eff}} + R_{\text{body}}$$

From the diode equation, we can express the voltage across the diode as the sum of the voltage drops across the series resistance and the voltage drop across the metallurgical junction [13]

$$I_{\text{D}} = I_{\text{s}}\left\{\exp\left(\frac{qV_{\text{D}} - \sum I_{\text{D}}R_i}{kT}\right) - 1\right\}$$

where the summation is of the voltage drops across the series resistances. The resistance can also be obtained from the derivative of the voltage with respect to current.

$$\frac{dV_D}{dI_D} = \frac{d(V_J + V_R)}{dI_D} = \frac{dV_J}{dI_D} + R_d$$

Solving for the resistance

$$R_d = \frac{dV_D}{dI_D} - \frac{dV_J}{dI_D}$$

From the high-level injection relationship

$$I_D = I_s \exp\left(\frac{qV_j}{2kT}\right)$$

and

$$\frac{dV_J}{dI_D} = \frac{2kT}{q}\frac{1}{I_D}$$

Substituting these expressions, we can solve for the diode series resistance term as

$$R_d = \frac{dV_D}{dI_D} - \frac{2kT}{q}\frac{1}{I_D}$$

Because of the self-heating, the resistance is a strong function of the temperature as the device is heated.

Figure 10.3 shows the ESD robustness of a SOI gated diode structure as a function of polysilicon perimeter. HBM results shows that the ESD results improve linearly with

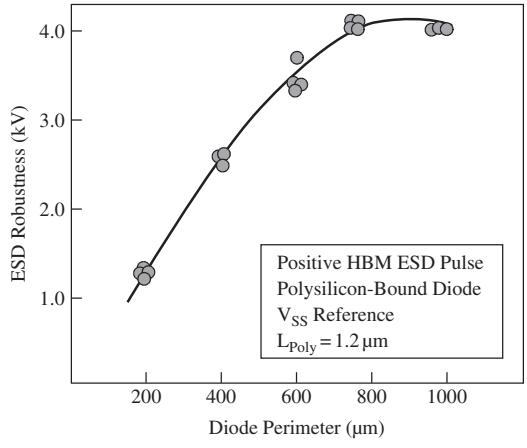

Figure 10.3 SOI polysilicon gated *p–n* diode (lubistor) HBM ESD results as a function of diode perimeter ($L_{poly} = 1.2\,\mu m$)

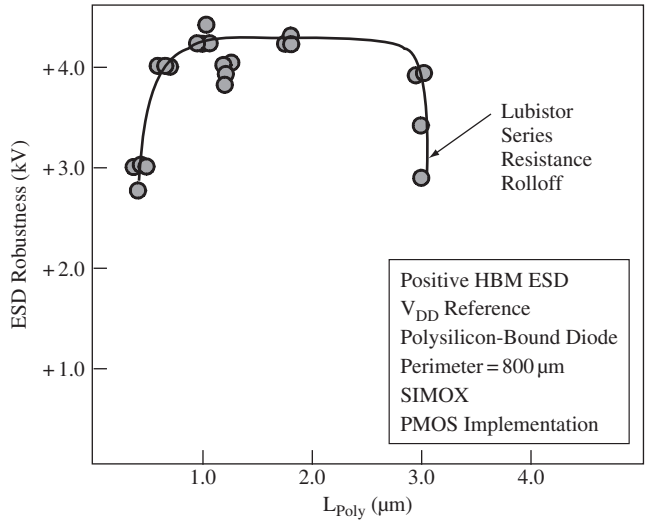

Figure 10.4 SOI polysilicon gated *p–n* diode (lubistor) HBM ESD results as a function of polysilicon gate length

increasing diode perimeter with a design L_{poly} of 1.2 μm. Figure 10.4 shows the HBM ESD results as a function of polysilicon length for an 800 μm perimeter diode structure. In this 0.25 μm technology, the ESD robustness of the diode structure is approximately 5 V/um (HBM) in prototype structures.

As the diode structure perimeter increases, the ESD robustness is well behaved. The linearity of the SOI diode structure indicates a lack of current constriction in a forward-bias mode of operation at high currents. The operation of the structure was not a strong function of the length of the center region over a wide range of operation. As the structure was reduced in length, the heavily doped p^+ implant and the n^+ implant spacing was reduced, leading to ESD failure. A flat zone where the ESD robustness was not a strong function of the length was then observed. As the length of the structure was increased, the ESD results began to decrease. This may be anticipated by the larger voltage drop across the anode–cathode region and failure due to increased Joule heating.

10.2.2 Extension Implant Technology

For SOI technologies based on a CMOS technology that uses extension implants, extension implants become part of the SOI ESD structure. In SOI technology, extension implants reduce the emitter and collector area for the lateral bipolar device; this reduces the bipolar current gain and improves short-channel effects. The source/drain junctions use a $TiSi_2$ salicide film to reduce the diffusion sheet resistance.

For an SOI lateral ESD diode element, the use of the extension implant provides a p^+ implant that extends under the gate region which is heavily doped. The p^+ emitter region in the extension implant extends into the cathode region beyond the p^+ source/drain implant, forming a nonuniform emitter profile. The cathode region also extends the

n^+ region into the n^- region, providing a nonuniform n^+/n^- profile. In this SOI diode structure, the extension implants lower the anode resistance (p^+ extension) and the cathode resistance (n^+ extension) providing a lower total diode series resistance. As a result, the SOI ESD model can be constructed as two parallel conductance terms that are summed to evaluate the total diode series resistance. The first conductance term represents the conductance near the gate interface, and the lower conductance addresses the lower implant regions.

$$R_{\text{anode}} = R_{\text{M}} + R_{\text{C}} + (R_{\text{Sal}})_{\text{eff}} + (R_{p+})_{\text{eff}}$$

$$R_{\text{cathode}} = R_{\text{M}} + R_{\text{C}} + (R_{\text{Sal}})_{\text{eff}} + (R_{n^+} + R_{n^-})_{\text{eff}}$$

In order to address the resistances, the parallel conductances along the surface and the lower region are integrated and must be considered.

In the experimental work performed in this technology generation, the technology supports a 5.0 nm oxide thickness. Figure 10.5 shows the SOI lateral gated diode with abrupt and extension implant drain structures. In this 0.20 μm technology, lateral polysilicon gated SOI diodes demonstrated an ESD (HBM) robustness level of 8.0 V/μm, which was superior to the ESD robustness of the equivalent structure in the 0.25 μm

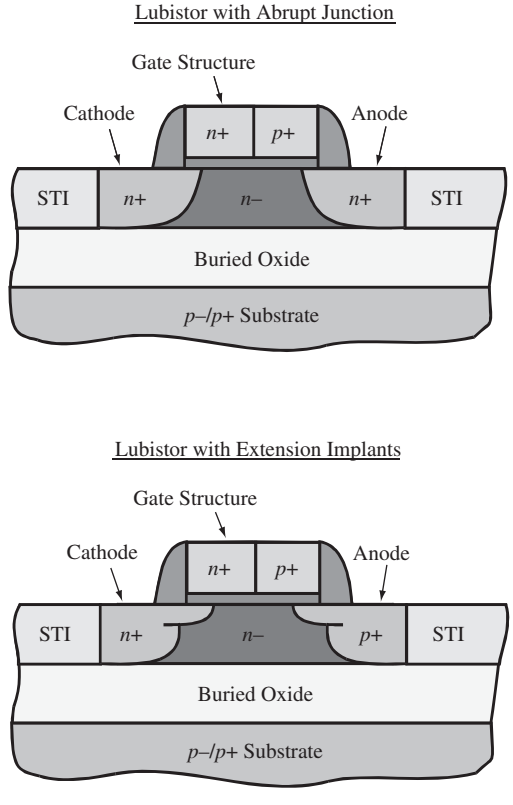

Figure 10.5 SOI lubistor with abrupt and extension implant source/drain structures

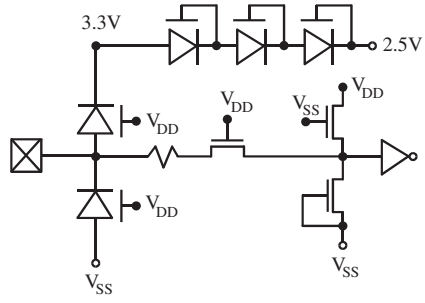

Figure 10.6 Circuit schematic of SOI ESD implementation

L_{eff} technology that used abrupt junctions. The extension implant dose was significantly higher than that used in the abrupt 0.25 μm L_{eff} technology [15–20].

As an example of its implementation, this ESD element was mapped into a bulk CMOS technology using STI isolation between the anode and cathode, and remapped into SOI using the SOI lateral gated diode ESD element [22]. The mixed-voltage application required 3.3 V tolerant I/O networks with a 2.5 V core voltage. In this design, the off-chip driver (OCD) networks were placed inside the center of the microprocessor chip, limiting the area allowed for the I/O network and ESD device. The C4 solder balls were connected to the receiver and OCD banks via aluminum interconnects. In the bulk implementation, the off-chip driver circuitry ESD results were limited by the aluminum interconnects. For the receiver networks, the ESD robustness of the receivers were limited by failure of the *n*-channel pass transistor and clamp networks. ESD results in the receiver networks were 4.3 kV (HBM) (Figure 10.6). This RISC-based microprocessor was mapped with no internal design changes. The ESD STI-bound diode networks were modified to polysilicon gated SOI lateral diode structures with identical area for the bulk and SOI ESD areas. The SOI ESD design area was divided so that the diode perimeter for the SOI diode to V_{DD} was the same as the one to the V_{SS} power rail. No additional area or masks were utilized for ESD protection. The off-chip driver circuit consisted of a self-biased well *p*-channel pull-up network and a cascoded series *n*-channel pull-down network. Resistor ballasting was employed in series with the n-channel pull-down network, using four 40Ω buried resistor (BR) elements. BR elements consist of a MOSFET gate structure, an n^+ source/drain implant for end contacts and an n^+ implant placed under the gate structure. The receiver networks comprised of an *n*-channel pass transistor, an *n*-channel grounded gate, *n*-channel clamp device and a *p*-channel MOSFET keeper element.

Experimental results of the bulk and SOI implementations are shown in Figure 10.7. No ESD failures were evident below 6.5 kV (HBM), whereas the bulk CMOS receiver failures occurred at 4.3 kV. From this first SOI implementation, it was clear that industry-acceptable ESD results were achievable in an advanced CMOS microprocessor without using additional design area, masks or implants.

10.2.3 Halo Implant Technology

For an SOI lateral ESD diode element, the element requires both usage of the p^+ MOFSET drain structure and the n^+ MOSFET drain structure where it is placed in

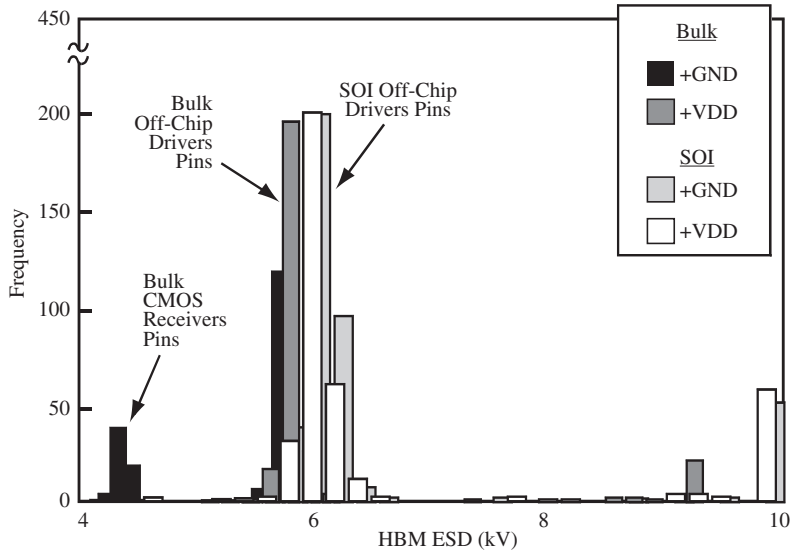

Figure 10.7 ESD HBM results of a 250 MHz RISC-based microprocessor in bulk CMOS and SOI technology using bulk p^+/n-well diodes and lateral SOI $p^+/n^-/n^+$ diodes, respectively.

either a *p*-type or *n*-type body. To prevent MOSFET punch-through, halo implants are used near the MOSFET junction regions to avoid extension of the depletion region and avoidance of MOSFET punch-through. For an *n*-channel MOSFET in a *p*-type well, a p^+ halo implant is used. For a *p*-channel MOSFET, a n^+ halo implant is used. When the SOI lateral gated diode element is formed a n^+ halo implant is placed in the body near the p^+ anode metallurgical junction, and a p^+ halo implant is placed near the n^+ heavily doped region in the body. In an implementation of this structure, since only one body is formed, one of the halos is of the opposite doping concentration to that of the SOI ESD diode body region. As a result, the 'bad halo' introduces undesired diode series resistance. To provide a good ESD SOI diode, this halo implant should be eliminated for good ESD diode operation.

10.3 SOI WITH INTERCONNECTS

10.3.1 Aluminum Interconnects

Interconnects and salicides can have a substantial influence on the ESD robustness in SOI technology as a result of the thermal dissipation through the contact metallurgy [15–20]. In an SOI chip implementation, alternative current paths are required which are available in the bulk implementation, but not the SOI designs. With the buried oxide film, the thermal dissipation through the contacts and the interconnects will play a larger role in the peak temperatures observed in the lateral SOI diode structures. In our structures, the physical length of the diode region is approximately the minimum channel length of the technology. With these relatively short diode resistor regions,

the anode and cathode resistances can be significantly reduced. As a result, interconnects and the salicide films will play a role in the dissipation of heat in the low-doped region of the SOI diode structure under the gate region. Additionally, ESD implementations in SOI technology can lead to alternative current paths which do not exist in an ESD implementation, but are evident in the bulk implemention.

As an example, a 330 MHz 1.8 V RISC-based microprocessor was mapped from bulk CMOS to SOI technology (Figure 10.8). The technology supported a 0.15 μm L_{eff} *n*-channel MOSFET and a 0.17 μm L_{eff} *p*-channel MOSFET device. The MOSFET drain used extension implants, halos and cobalt salicide ($CoSi_2$) junctions. The source/drain extensions provided good V_T rolloff characteristics and the $CoSi_2$ allowed the ability to maintain low-salicided polysilicon gate structure sheet resistance.

The semiconductor chip, in the bulk CMOS implementation, did not have ESD protection limited by the interconnect failure. In the bulk implementation, the ESD network comprised a single diode STI-bound p^+/n-well diode to V_{DD} and a single *n*-well-to-substrate diode element to V_{SS}. ESD results in bulk CMOS chip showed no fails below 10 kV HBM levels. In the SOI implementation, a 700 μm perimeter SOI $p^+/n^-/n^+$ polysilicon gated diode ESD network was connected to V_{DD} and to V_{SS}. The ESD area was allocated equally for the ESD diode to V_{DD} and V_{SS}. Worst-case ESD failures first occurred at 9.4 kV relative to the V_{SS} power rail for positive polarity. HBM testing relative to the V_{DD} power rail showed no failures below 10 kV HBM levels for the positive polarity. For negative pulse testing, no ESD failures occurred below −10 kV relative to the V_{SS} ground rail. Worst-case negative pulse failures occurred at −7.0 kV relative to V_{DD} power rail and −7.2 kV relative to V_{DD}.

Failure analysis was completed on the worst-case failure mechanisms. Failure analysis results show that the aluminum interconnects in the power bussing was the ESD-limiting failure mechanism in this design.

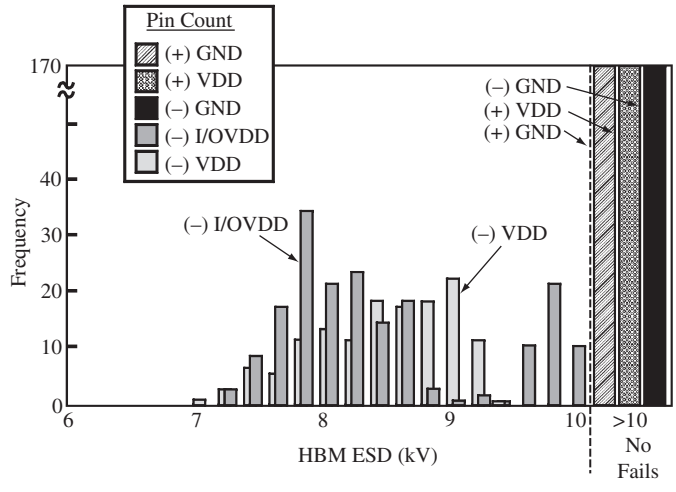

Figure 10.8 ESD HBM results of a 330 MHz 1.8 V SOI microprocessor in a 0.15 μm L_{eff} technology

10.3.2 SOI and Copper Interconnects

With the introduction of copper interconnects into SOI technology, the ability to lower the thermal resistance of the SOI diode structure can be further improved. In the 0.12 μm L_{eff} technology generation, Cu interconnects with SiO_2 interlevel dielectrics are introduced to provide reduced interconnect RC delay. As discussed in Chapter 9, the ESD robustness of the Cu interconnects are superior to the Al interconnects. It has been shown that Cu interconnects achieve a two-fold improvement in the critical current density to failure J_{crit}. In this technology, there are six levels of Cu interconnects, a 0.12 μm L_{eff} *n*-channel MOSFET, and a 0.15 μm L_{eff} *p*-channel MOSFET (where both transistors have extension implants, $CoSi_2$, and a 3.5 nm dielectric thickness).

Additionally, from the thermal transmission line model, the transmission line equation can be represented as a thermal analog

$$\frac{\partial^2 T(x,t)}{\partial x^2} - RY(t)T(x,t) = Rg(t)$$

The admittance Y(t) is associated with the thermal network consisting of the vertical stratified medium which is in parallel to the interconnects and the interlevel dielectric films.

$$Y(t) = \sum_i Y(t)$$

With the introduction of MO wiring and copper interconnects, the thermal admittance within the ESD elements can be further reduced. Using wide Cu plates, the thermal admittance can be lowered reducing the peak temperature in the ESD diode structures.

A multi-finger lateral SOI $p^+/n^-/n^+$ diode structure consisting of a two-finger, four-finger and seven-finger diode structure where each finger has a perimeter of 68 μm, was constructed. The SOI ESD network has local W 'M0' bar interconnects placed along the complete length of the p^+ and n^+ contact regions. This provides low electrical and thermal resistance on the diffusion areas. The local M0 W interconnect is contacted with W contacts and connected to the first-level M1 Ti/Al/Ti metallurgy. The M1 interconnects are then connected to the M2 copper interconnects. Figure 10.9 shows

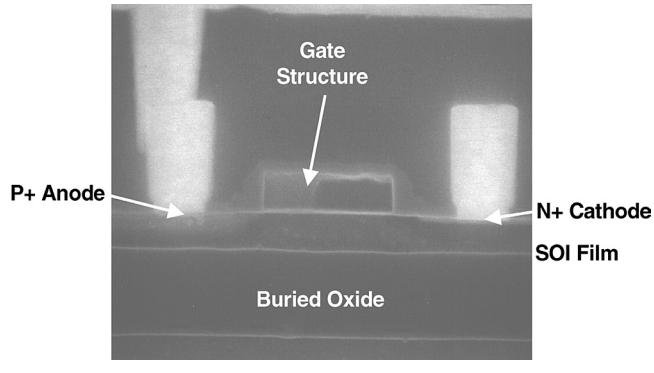

Figure 10.9 Cross-section of an SOI ESD structure

an enlargement of the SOI ESD structure. The polysilicon ring forms the gate structure with the gate contact landing over the STI isolation, as can be seen in Figure 10.10.

Figure 10.11 shows the ESD results as a function of ESD perimeter. HBM ESD results of 4 kV are achieved with the 169 μm two-finger diode structure. ESD results increased with increasing p^+ anode perimeter for both the 338 and 591 μm structures.

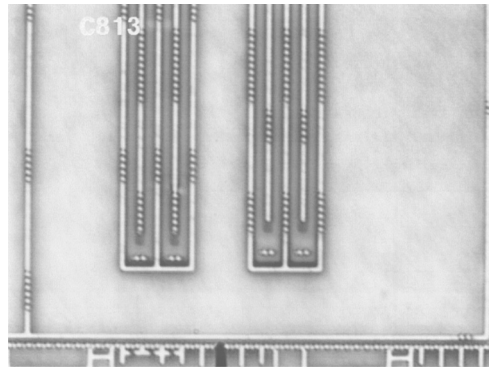

Figure 10.10 SOI ESD lubistor double-diode implementation

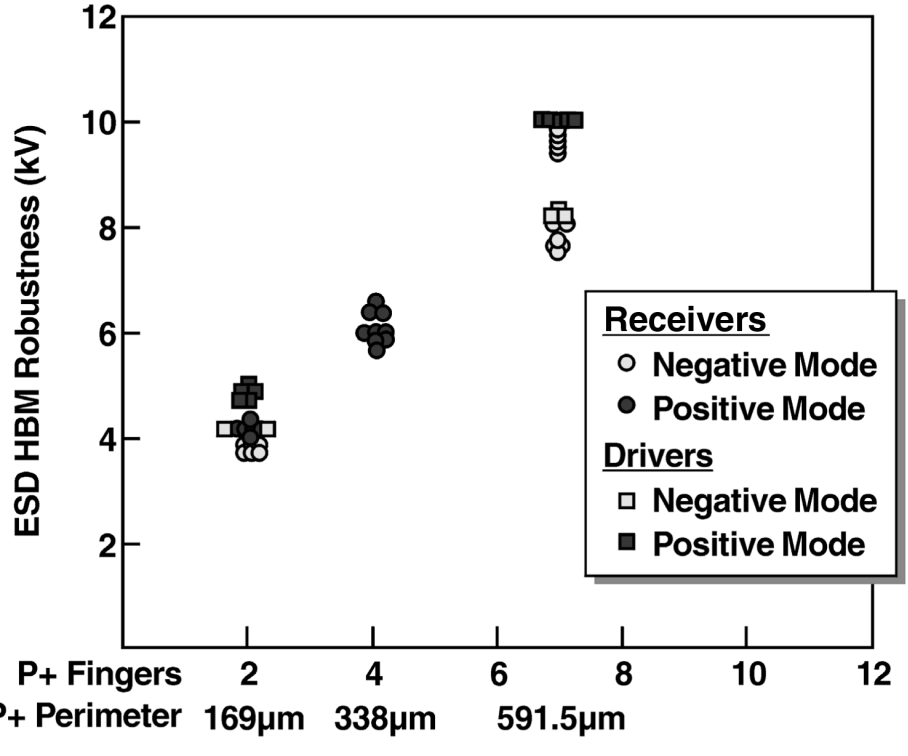

Figure 10.11 HBM ESD results of SOI receiver and driver circuits as a function of SOI ESD lubistor perimeter

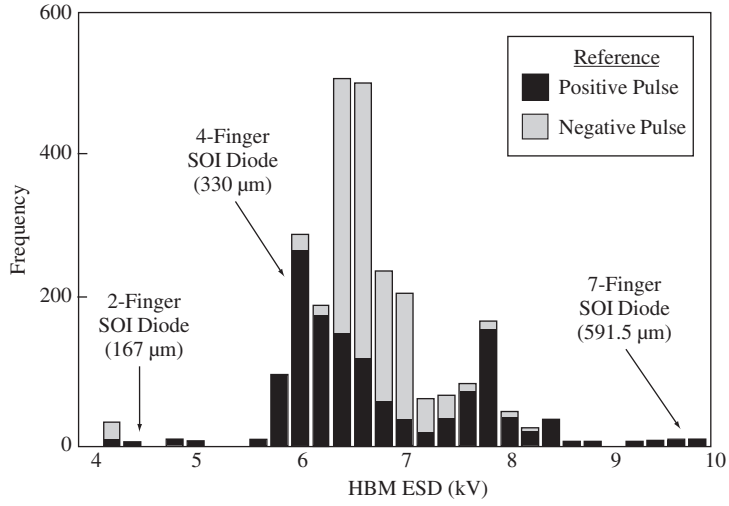

Figure 10.12 Histogram of SOI ESD results as a function of SOI ESD lubistor perimeter

ESD results higher than 8 kV were achieved using the 591 μm diode structure. Failure analysis demonstrated in all three ESD cases that the ESD failure mechanism occurred in the ESD network between the p^+ anode and n^+ cathode. From this work, an equivalency of 23.8 V/μm diode is achieved with the smallest two-finger ESD design, and the seven-finger structure shows an equivalency of 14.2 V/μm (the actual scaling slope shows 11 V/μm scaling) [18–20].

Figure 10.12 shows the ESD distribution for the three different sized ESD designs. The I/O design was optimized to trade off ESD robustness against the capacitive loading effect of the ESD network, where it was decided to use the four-finger SOI structure, which provided minimal capacitance loading effects yet maintained ESD robustness well over 4 kV.

A key observation from this experimental work can be understood. Since SOI technology has no substrate connected to the various fingers of the SOI ESD structure, the scaling of the ESD level increases linearly with the number of fingers and the design perimeter. This indicates also that there is little evidence of nonuniform current flow in the SOI diode structure. What is evident is the failure mechanism occurs between the minimum distance between an anode contact and a cathode contact. Hence, many of the concerns that exist in bulk CMOS because of nonuniform current distribution in wells and substrates do not appear to be significant. The elimination of the substrate and well regions leads to a decoupling of the devices improving current uniformity.

10.3.3 Scaling

From the ESD testing of the SOI polysilicon gated diodes in successive technology generations, it is clear that the technology migration choices and scaling of the technology have led to improved ESD robustness in the SOI lateral gated $p+/n^-/n+$ diode structure [15–20]. These include the transition from $TiSi_2$ to $CoSi_2$, Al to Cu interconnects and

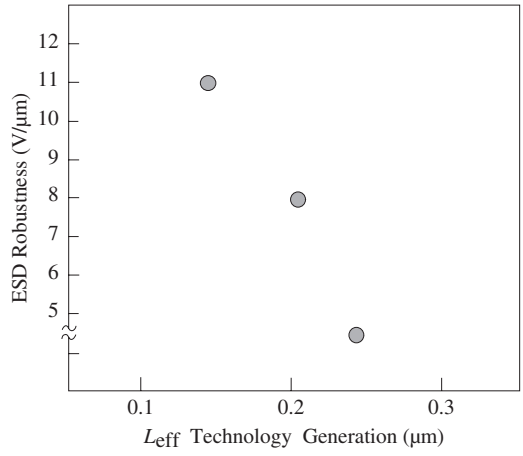

Figure 10.13 Normalized HBM ESD robustness as a function of the MOSFET channel length technology generation

dimensional scaling. Figure 10.13 shows the ESD robustness of the SOI diode network as the structure migrated from the 0.25 to the 0.12 μm technology generation [16–20]. In the 0.25 μm technology generation, ESD robustness of 5 V/μm was demonstrated. ESD results of 4 kV was achieved in the test site structure with 800 μm ESD structures. As the designs were migrated to 0.12 μm L_{eff} SOI technology, ESD results continue to show an improvement. In the 0.12 μm L_{eff} SOI technology, ESD results over 8 kV are demonstrated in high-pin-count microprocessor chips with structures smaller in design perimeter and in physical area. At this point, our data provides evidence of improvement in the ESD robustness of the ESD network with technology migration and scaling of the base CMOS technology from the 0.25 to the 0.12 μm technology generation.

As SOI technology is scaled, additional structural scaling will be addressed. As the technology is scaled, the buried oxide (BOX) region will also be scaled. The scaling of the BOX region will reduce the thermal resistance to the substrate. As the BOX thickness is scaled, the thermal resistance to the MOSFET channel region and the ESD structures will be reduced. Additionally, as the technology is scaled the silicon film will also be scaled. Today, in ultra-thin SOI (UTSOI), ESD protection is being practiced using SOI lateral $p^+/n^-/n^+$ diode structures. As the silicon film is scaled the ESD protection will become more difficult.

10.4 ALTERNATIVE DEVICES

10.4.1 SOI Diodes with Both Body Dopant Types

For an SOI lateral ESD diode element, the element requires use of both the p^+ MOSFET drain structure and the n^+ MOSFET drain structure where it is placed in either a *p*-type or *n*-type body. An alternative implementation can be used where both dopant types exist in the body region. In this case a $p^+/p^-/n^-/n^+$ SOI lateral diode ESD structure is formed. In the prior structure, the metallurgical junction at the $p+/n^-$

interface is a one-sided metallurgical junction since the p^+ region is heavily doped. In the $p^+/p^-/n^-/n^+$ SOI ESD element, the metallurgical junction is a two-sided junction, allowing the depletion region to extend into both the anode and the cathode regions. Experimental results by Ker *et al.* demonstrated improved results compared with the three-region $p^+/n^-/n^+$ SOI lateral gated diode 2 [23].

10.4.2 Ungated SOI Diode Structure

For an SOI lateral ESD diode element, the ESD element has a MOSFET gate structure used as a masking structure for drain implant formation and as a means of separating the anode and cathode salicide film region (e.g. to prevent shorting). An advantage of ungated ESD elements is dielectric overstress. In these structures, the SOI ESD MOSFET gate structure can lead to HBM or CDM failures, depending on where the gate structure is connected. There are two solutions to electrical overstress of the gate structure. To improve ESD robustness of the SOI ESD element, Ker *et al.* physically removed the MOSFET gate structure, using an etch process to form an ungated SOI ESD diode element [23]. SOI $p^+/p^-/n^-/n^+$ lateral ESD elements were constructed without the gate structure, achieving improved ESD results. Alternative approaches, such as circuit solutions, can also be utilized.

10.5 SOI DYNAMIC THRESHOLD MOSFET AND ESD

SOI has a unique advantage over single- and double-well CMOS technology because of its ability to separate the channel region from the substrate region. This allows the ability to bias the channel region of both the *p*- and *n*-channel transistors independently from the substrate or well regions. In this section, we will introduce the idea of dynamic threshold MOSFETs (DTMOS) and apply the concepts for ESD protection elements [22,24–26].

SOI MOSFETs, in functional and ESD applications, are significantly different from standard bulk MOSFET structures. SOI transistor operation is dependent on the state of the SOI MOSFET 'body' region. The SOI MOSFET body potential is dependent on the capacitive coupling to the SOI MOSFET drain, source, gate, and substrate regions. When voltages are applied to the gate region, competition between the capacitance elements will determine the eventual potential of the body region. The potential of the body modulates the MOSFET threshold voltage needed on the MOSFET gate to invert the surface and turn on the MOSFET device. In a bulk CMOS device, the MOSFET body effect is a function of the well or substrate bias. These can not be independently biased as they are coupled to other circuit elements. The MOSFET threshold voltage is the sum of the interface potential jump ϕ_{MS}, the Fermi potential ϕ_F, and the voltage drop across the oxide and silicon [24,25].

$$V_T = \Phi_{MS} + \Phi_F - \frac{Q_{ox}}{C_{ox}} + \frac{Q_b}{C_{ox}}$$

where

$$Q_b = \sqrt{2\varepsilon_{Si} q N_A (2\Phi_F - V_B)}$$

where V_B is the body voltage. The threshold voltage can be rewritten as

$$V_T(V_B) = V_{TO} + \gamma(\sqrt{2\Phi_F - V_B} - \sqrt{2\Phi_F})$$

where

$$V_{TO} = \Phi_{MS} + \Phi_F - \frac{Q_{ox}}{C_{ox}} + \gamma\sqrt{2\Phi_F}$$

and

$$\gamma = \frac{\sqrt{2\varepsilon_{Si} q N_A}}{C_{ox}}$$

SOI technology has an advantage over twin-well bulk CMOS in that the *n*-channel MOSFET substrate is not bound to the substrate potential. SOI technology has the advantage of allowing the *n*-channel and *p*-channel MOSFET body to float, allowing the body to couple to the drain node, and be electrically disconnected from the V_{SS} and V_{DD} power rails. This can be advantageous for ESD protection networks. Partially depleted (PD) SOI technology, with body contacts, allows for ESD circuits which can use complementary methods for the *n*-channel and *p*-channel transistor to address positive and negative ESD pulse events.

A first advantage of the DTMOS device for ESD protection is the use of the body potential to provide a low trigger ESD device. When the body voltage V_B is increased, the SOI DTMOS threshold decreases. When the DTMOS device is used as a trigger element, as the body voltage increases, the trigger voltage decreases.

A second advantage of DTMOS devices is that the SOI MOSFET current drive is higher when the gate is activated. As the MOSFET body voltage increases, the DTMOS current drive increases. From the current equation of the MOSFET devices[24,25].

$$I_{dsat} = \mu C_{ox} \frac{W}{L}[V_g - V_T(V_B) - V_{dsat}]$$

where

$$V_{dsat} = \frac{E_{sat} L[V_g - V_T(V_B)]}{E_{sat} L + [V_g - V_T(V_B)]}$$

The saturated transconductance g_{msat} is the derivative of the drain saturation current with respect to the gate voltage. From the above relationships, as the body voltage increases, the SOI DTMOS threshold voltage decreases, leading to a higher drain saturation current. As a result, the saturated transconductance of a DTMOS device increases above the standard SOI MOSFET transistor.

10.5.1 SOI Dynamic Threshold ESD Structures

For integration of the dynamic threshold concepts into an ESD element, two classes of ESD network can be constructed. The DTMOS device can be used in both a transistor or diode mode of operation. To use the SOI DTMOS device as a transistor, the source and drain electrodes are separated, whereas the body and gate are coupled together to provide the DTMOS effect as the gate voltage is biased.

In a second mode of operation, the DTMOS device can be used in a diodic mode of operation. In this case, the body, drain and gate are connected to one electrode and the source to a second electrode. In this fashion, body-coupling, and gate-coupling are integrated with the drain to provide a body-, gate- and drain-coupled ESD network, which we will refer to as a body/gate-coupled DTMOS SOI diode element. In this fashion, the element has two parallel operations within the structure. The device acts as a SOI lateral gated diode element (e.g. $p^+/p^-/n^+$) where the p^+ is the SOI MOSFET body contact, the p^- region is the channel or body region of the SOI transistor, and the n^+ region is the source of the SOI transistor. This is in parallel to the SOI DTMOS element. As a result the structure has both a diodic and transistor mode of operation where conduction occurs in the MOSFET body region in the SOI lateral $p^+/p^-/n^+$ region of the structure, and surface conduction in the MOSFET region of the body. By evaluation of the current drive of the element, it can be shown that the current drive increases as a result of the bipolar conduction and the MOSFET surface conduction.

Figure 10.15 shows an example of implementation of a body/gate-coupled DTMOS SOI diode structure. Three different structures can be constructed where in all cases a polysilicon ring is formed about the n^+ drain, and the polysilicon gate structure encloses the body contact and the n^+ drain (Figure 10.14).

The n^+ MOSFET source encloses the polysilicon ring gate structure. In all these structures, the p^+ body contact (at the drain side) abuts the polysilicon ring serves as a body contact for the MOSFET and forms a lateral SOI gated p–n diode structure (SOI lubistor) adjacent to the MOSFET structure. In the first design, the p^+ body contact abuts the n^+ drain. In this implementation, it was believed that the butted structure would provide the lowest dynamic resistance, and the most space efficient. The butted

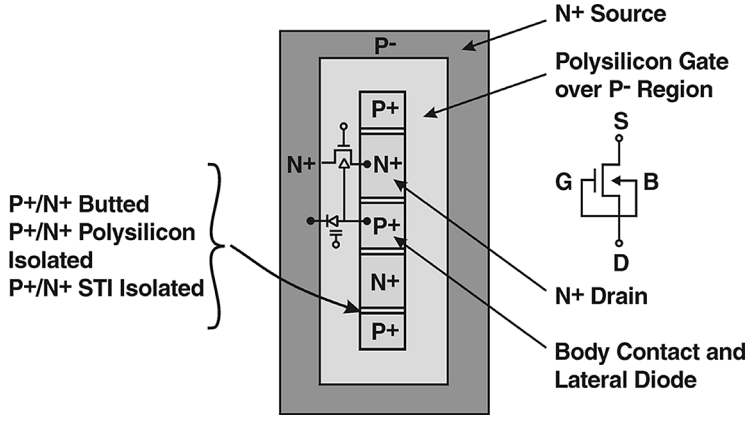

Figure 10.14 Layout of dynamic threshold body/gate-coupled SOI diode structure

structure is also bridged by cobalt salicide. In the second design of the DTMOS device, the p^+ body contact is separated from the n^+ drain region. This design will avoid any technology-related concerns of butted structures and allows for shallow trench isolation (STI) of the body-contacted region from the MOSFET drain region. In the third structure, the p^+ body contact is separated by a polysilicon gate structure from the n^+ drain. The third implementation provides polysilicon isolation of the body and the MOSFET drain region, avoiding any STI pull-down mechanisms, independent of biasing capability and introducing an additional lateral diode between the body and the source.

Figure 10.15 shows TLP results of an SOI off-chip driver (OCD) network and the three different large dynamic threshold body/gate-coupled ESD diode networks. A total of 24 different 10-finger DTMOS ESD networks were tested to evaluate the ability to protect an SOI driver circuit.

In the measurements, the body/gate-coupled DTMOS ESD device is configured with the drain, gate and body connected to the input node and the source connected to V_{DD} power supply. The results show that the SOI body/gate-coupled DTMOS ESD network with STI between the p^+ and n^+ region provides a higher R_{ON} relative to the other two designs (e.g. $R_{ON} \simeq 4\,\Omega$). This can be understood in that the trench isolation does not provide any means of conduction and causes more current crowding and resistance. In the second implementation, where the polysilicon gate region is used between the body and the drain, the addition of the extra diode region provides a lower lateral body resistance and a lower R_{ON} in the ESD network ($R_{ON} \simeq 2.8\,\Omega$). In the third implementation, where the p^+ body and n^+ drain are abutted, $R_{ON} \simeq 1.2$.

To better understand the operation of these structures, it is important to evaluate a matrix where the ratio of the MOSFET drain width and the body width are varied while the total length of the structure remains fixed. In our experimental matrix, as the body

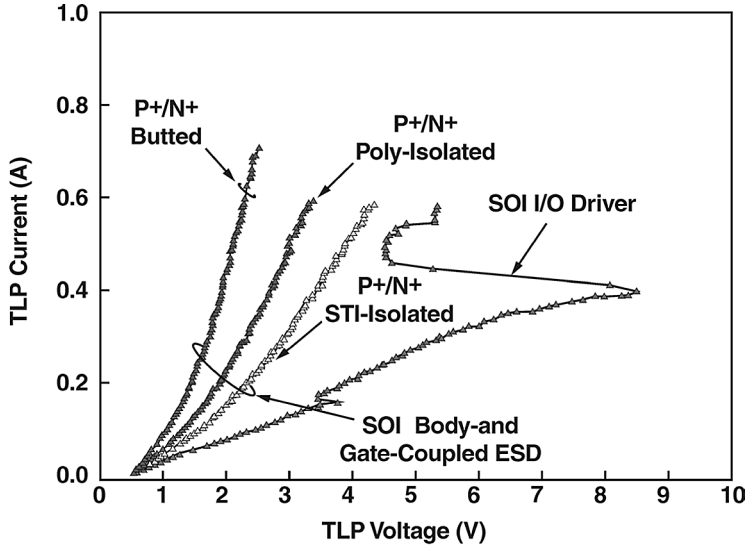

Figure 10.15 TLP measurements of DTMOS body/gate-coupled SOI diode structure for three different implementations

width was increased, the MOSFET width was decreased so that the total perimeter of the source is the same width. Figure 10.16 shows an example of the body/gate-coupled DTMOS ESD network, where the ratio of the body and drain width are varied. Defining the body contact width W_{BC} and the drain width W_N we can form a ratio, or percentage body contact

$$\Psi = \frac{W_{BC}}{W_{BC} + W_N}$$

where the total body contact width is

$$(W_B)_T = \Psi W$$

and the total DTMOS transistor width is

$$(W_{DTMOS})_T = (1 - \Psi)W$$

where W is the total width of the ESD structure. From this structure, the total current can be predicted from the diode equation, and the DTMOS drain-to-source saturation current

$$I_D = I_s \Psi W \left\{ \exp\left(\frac{qV_d - \sum I_D R_i}{kT} \right) - 1 \right\}$$

$$I_{dsat} = \mu C_{ox} \frac{W}{L} (1 - \Psi)[V_g - V_T(V_B = V_d) - V_{dsat}]$$

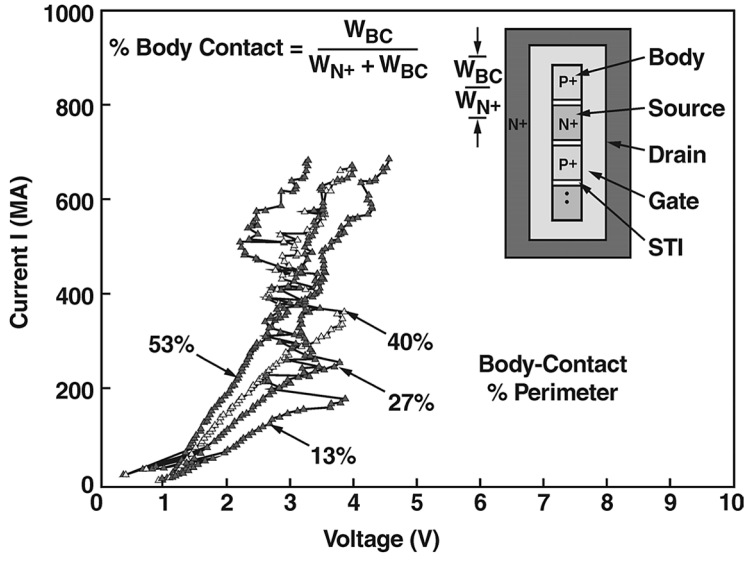

Figure 10.16 TLP measurements of DTMOS body/gate-coupled SOI diode with different body contact-to-MOSFET ratio

In the first equation for the diode, the saturation current is normalized per unit width of the device region, and the voltage on the diode is the drain voltage. In the DTMOS device, the body and gate voltage are equal to the drain voltage. These equations will be valid until the device undergoes MOSFET snapback, where the avalanche physics terms must be added to the device equations.

In Figure 10.16, TLP *I–V* measurements are taken for devices as a function of the percentage of body contact (e.g. for the cases of 13%, 27%, 40% and 53% body contact). Observation of the 13% body contact, the first TLP *I–V* characteristic shows that it follows a monotonically increasing current from 0 to 4 V, and this is followed by snapback of the DTMOS device. As the body voltage rises, the diode formed between the body contact and the MOSFET source first turns on, and at the same time, the MOSFET threshold voltage decreases. As the gate voltage exceeds the threshold voltage, the dynamic threshold MOSFET device also turns on, with a high I_{dsat}. When the voltage across the structure approaches the snapback voltage, this structure undergoes a snapback state. As the percentage body contact ratio increases, the dynamic on-resistance increases. With lower on-resistance, the discharge current capability improves.

10.6 SOI AND ESD IN THE FUTURE

As SOI is scaled to ultra-thin SOI, the ability to achieve good ESD results will become a greater challenge. As SOI scales, the buried oxide film will be scaled to reduce processing time which improves the thermal resistance to the substrate. Additionally, new structures, such as double-gate SOI, and FINFET structures will begin to emerge as new SOI structures. The presence of the second gate will improve the current drive and lower the thermal resistance of the SOI structure. Continued work will be needed to maintain ESD robust structures in future SOI technologies.

In Chapter 11, silicon–germanium (SiGe) and silicon–germanium–carbon (SiGeC) technologies and ESD phenomena will be discussed. The chapter will focus on the distinctions between Si bipolar versus SiGe, and the SiGe versus SiGeC devices from an ESD perspective. The first paper on ESD in SiGe bipolar transistors was published as recently as 2000. Someday the material in Chapters 10 and 11 will be connected as these technologies are integrated. For now, let's look at SiGe and ESD.

PROBLEMS

10.1. Derive the thermal resistance of a silicon-on-insulator wafer which consists of a thin silicon film t_{film}, a buried oxide film t_{ox}, and substrate t_{sub}.

10.2. Derive the thermal capacitance of a silicon-on-insulator wafer which consists of a thin silicon film t_{film}, a buried oxide film t_{ox}, and substrate t_{sub}.

10.3. Construct an electrical equivalent model using the thermal resistor and thermal capacitance.

10.4. Treating the SOI wafer as a stratified media, solve for the temperature of each domain at the boundaries and the center using the transfer matrix approach, assuming a known wafer temperature at ambient.

10.5. Derive a relationship between the power to failure of an SOI transistor and the maximum power dissipation during circuit operation.
10.6. Derive an electrical model of a lateral gated diode structure assuming a $p^+/n^-/n^+$ diode (e.g. whose body region is n^-).
10.7. Derive an electrical model of a lateral gated diode structure assuming a $p^+/p^-/n^+$ diode (e.g. whose body region is p^-).
10.8. Derive an electrical model of a lateral gated diode structure assuming a $p^+/p^-/n^-/n^+$ diode (e.g. whose body region is p^- and n^- of a given defined length).
10.9. Derive the power to failure of an SOI transistor compared with a bulk CMOS transistor based on the Wunsch–Bell model. Derive it according to the Dwyer model.
10.10. Given that an SOI transistor buried oxide region is scaled for each technology generation by the MOSFET constant electric field scaling parameter α, how does the total thermal resistance scale with technology scaling assuming the thin silicon film is fixed in dimension ? Assuming the thin film is scaled with the buried oxide film, how does the total thermal resistance scale ?
10.11. Derive the ratio of the current drive of a dynamic threshold MOSFET to the current drive of a standard MOSFET. Derive the ratio of the turn-on voltages. Derive the required width of a standard MOSFET necessary to achieve the same current drive as a DTMOS device of width W.
10.12. Derive a model of dynamic threshold MOSFET SOI ESD device with a parallel diode body element as shown in this chapter, where the total width of the device is the sum of the diode width and the MOSFET width.

REFERENCES

1. C. Hu. Low voltage CMOS device scaling. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, February 1994; 86–87.
2. G. Shahidi *et al.* CMOS scaling in the 0.1-μm 1.X V regime for high performance applications. *IBM Journal of Research and Development* 1995; **39**: 229–244.
3. J. P. Colinge. *Silicon-on-Insulator Technology: Materials to VLSI*, Dordrecht: Kluwer, 1991.
4. G. Shahidi *et al.* Partially depleted SOI technology for digital logic. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, Session 25, WP 25.1, February 1999; 426–427.
5. M. Canada *et al.* A 580 MHz RISC microprocessor in SOI. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, Session 25, WP 25.3, February 1999; 430–431.
6. Y. Kim. A 0.25 μm 600 MHz 1.5 V SOI 64b ALPHA Microprocessor. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, Session 25, WP 25.4, February 1999; 432–433.
7. D. Allen. A 0.2 μm 1.8 V SOI 550 MHz 64b powerPC microprocessor with copper interconnects. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, Session 25, WP 25.7, February 1999; 438–439.
8. M. Chan, S. S. Yuen, Z. J. Ma, K. Y. Hui, P. Ko and C. Hu. Comparison of ESD protection capability of SOI and bulk CMOS output buffers. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1994; 292–298.
9. K. Verhaege, G. Groesenekan, J. P. Colinge, H. Maes. Analysis of snapback in SOI MOSFETs and its use for an SOI ESD protection circuit. *Proceedings of the IEEE SOI Conference*, 1992; 140–141.

10. K. Verhaege, G. Groseneken, J. P. Colinge and H. Maes. The ESD protection of SOI snapback MOSFETs: mechanisms and failure modes. *Proceedings of the EOS/ESD Symposium*, 1993; 215–219.
11. R. Joy and E. S. Schlig. Thermal properties of very fast transistors. *IEEE Transactions on Electron Devices*, 1970; **ED-17**, (8): 586–593.
12. P. Raha, S. Ramaswamy and E. Rosenbaum. Heat flow analysis for EOS/ESD protection device design in SOI technology. *IEEE Transactions on Electron Devices*, 1997; **ED-44** (3): 464–471.
13. Y. Wang, P. Juliano and E. Rosenbaum. Electrothermal modeling of ESD diodes in bulk-Si and SOI technologies. *Proceedings of the EOS/ESD Symposium*, 2000; 430–437.
14. S. Voldman. Semiconductor diode with silicide films and trench isolation. U.S. Patent 5,629,544, May 15, 1990.
15. S. Voldman *et al*. CMOS-on-SOI ESD protection networks. *Proceedings of the EOS/ESD Symposium*, 1996; 291–301.
16. S. Voldman *et al*. CMOS-on-SOI ESD protection networks. *Journal of Electrostatics* 1998; **42**, 333–350.
17. S. Voldman. The impact of MOSFET technology evolution and scaling on electrostatic discharge protection. *Microelectronics Reliability* 1998; **38**: 1649–1668.
18. S. Voldman. The impact of technology evolution and scaling on electrostatic discharge (ESD) protection in high-pin-count high-performance microprocessors. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, Session 21, WA 21.4, February 1999; 366–367.
19. S. Voldman *et al*. Electrostatic discharge (ESD) protection in silicon-on-insulator (SOI) CMOS technology with aluminum and copper interconnects in advanced microprocessor semiconductor chips. *Proceedings of the EOS/ESD Symposium*, 1999; 105–115.
20. S. Voldman, D. Hui, L. Warriner, D. Young, R. Williams, J. Howard, V. Gross, W. Rausch, E. Leobangdung, M. Sherony, N. Rohrer, C. Akrout, F. Assaderaghi and G. Shahidi. Electrostatic discharge protection in silicon-on-insulator technology. *Proceedings of the IEEE International SOI Conference*, Session 5.1, 1999; 68–72.
21. S. Voldman *et al*. Semiconductor process and structural optimization of shallow trench isolation-defined and polysilicon-bound source/drain diodes for ESD networks. *Proceedings of the EOS/ESD Symposium*, 1998; 151–160.
22. S. Voldman, D. Hui, D. Young, R. Williams, D. Dreps, M. Sherony, F. Assaderaghi and G. Shahidi. Silicon-on-insulator dynamic threshold ESD networks and active clamp circuitry. *Proceedings of the EOS/ESD Symposium*, 2000; 26–28.
23. M. D. Ker, K. K. Hung, H. T. H. Tang, S. C. Huang, S. S. Chen and M. C. Wang. Novel diode structures and ESD protection circuits in a 1.8 V 0.15 μm partially depleted SOI salicided CMOS process. *Proceedings of the International Physical and Failure Analysis (IPFA) of Integrated Circuits Symposium*, Singapore, 2001; 91–97.
24. F. Assaderaghi *et al*. A dynamic threshold voltage MOSFET (DTMOS) for ultra-low voltage operation. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 1994; 809–812.
25. S. Voldman, F. Assaderaghi, J. Mandelman, L. Hsu and G. Shahidi. Dynamic threshold body and gate-coupled SOI ESD protection networks. *Proceeding of the EOS/ESD Symposium*, 1997; 210–220. *Journal of Electrostatics*, 1998; **44**: 239–255.
26. F. Assaderaghi, J. Mandelman, L. Hsu, G. Shahidi and S. Voldman. Silicon on insulator body-coupled gated diode for electrostatic discharge (ESD) and analog applications. U.S. Patent 5,811,857, September 1998.

11 Silicon–Germanium and ESD

In this chapter, we will discuss silicon–germanium (SiGe) and ESD. The first paper on ESD in SiGe bipolar transistors was published as recently as 2000. This chapter will bring the reader up to speed from the first publication to the present. The first electrical measurements of SiGe HBT devices will be shown, as well as the first SiGe–carbon (SiGeC) measurements. The chapter will focus on the distinctions between Si bipolar homojunctions versus SiGe heterojunction bipolar transistors (HBT), SiGe versus SiGeC HBT, and other aspects from an ESD perspective.

In the 1950s and 1960s, the question of the limitations of the transistor speed was being addressed by device researchers such as Early, Johnson, Kirk and others [1–7]. The work by H. Kroemer on the heterojunction opened a new door to achieving higher speeds than anticipated with silicon homojunctions [8,9]. Significant growth in both high-frequency wired and wireless markets has introduced new opportunities where compound semiconductors have unique application advantages over bulk CMOS technology [10]. With the rapid advancement of the epitaxial layer pseudomorphic silicon–germanium (SiGe) deposition processes, epitaxial–base SiGe heterojunction bipolar transistors (HBT) have been integrated with mainstream advanced CMOS development processes, providing the advantages of SiGe technology for analog and RF circuitry while maintaining the full utilization of the advanced CMOS technology base for digital logic circuitry [11–23].

Significant work has been completed to understand the high-frequency, and high-current effects of the SiGe HBT device [24–31]. Although significant work was done to understand the ESD robustness of the homojunction transistor [32–44], no publications appeared on the ESD robustness of silicon–germanium transistors, although they were rapidly gaining interest as mainstream semiconductor components. SiGe HBT devices are replacing silicon bipolar junction (BJT) devices as the primary element in all analog and RF applications. With the increased volume and growth in the applications that use SiGe HBT devices for external circuitry, the quantification of the SiGe HBT devices during electrostatic overstress (EOS) and electrostatic discharge (ESD) sensitivities has become more importance [45–55].

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

11.1 SILICON–GERMANIUM

11.1.1 Silicon Germanium Structures

Epitaxial deposition of silicon germanium allows for optimization of a bipolar transistor without the limitations and pitfalls of implanted dopants. The SiGe alloy film, which can in principle be an atomically abrupt interface, avoids limitations of Gaussian profiles, dopant implant channelling and hot-process deficiencies. The ultra-high vacuum chemical vapor deposition (UHV/CVD) allows for tight control of dopant implants and low thermal budget. Figure 11.1 shows a cross-section of an epitaxial graded Ge-base SiGe HBT device. The SiGe HBT device is formed on a *p*-silicon substrate wafer. This is to provide good noise isolation between the analog and digital circuitry and a low substrate-to-sub-collector capacitance. The polysilicon deep trench isolation defines the sub-collector region and also provides a low sidewall capacitance. Additionally, the deep trench isolation influences the subcollector-to-substrate breakdown voltage and eliminates parasitic devices in a SiGe HBT device. The polysilicon deep trench isolation is used for high-performance applications as it provides a lower unity power cutoff frequency f_{MAX}. For low-cost processes, the deep trench is replaced with either a low-cost trench isolation (TI) or diffusion implants. The DT-defined collector contains a highly doped n^{++} subcollector, n^{+} collector reach-through implant and optional n^{++} pedestal implant. The pedestal implant, used for the high-frequency SiGe *npn*, provides a low-resistance collector, a high f_T, and improved Kirk effect. From the CMOS base technology, the shallow trench isolation (STI) is used for definition of the collector contact and sub-collector reach-through structure. Silicon–germanium transistors can also be constructed with LOCOS isolation from older generation CMOS or BiCMOS base technology. The deposition of the SiGe film has been demonstrated in

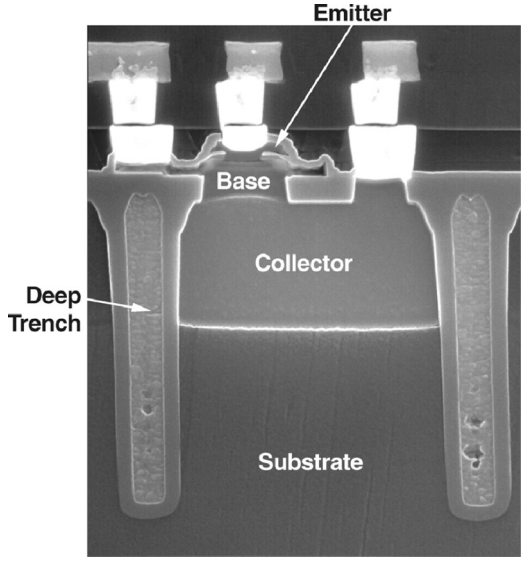

Figure 11.1 SiGe heterojunction bipolar transitor (HBT) device

a number of different chemical vapor depositions. Using UHV/CVD processing, SiGe is deposited in the base region over single-crystal silicon and the STI isolation structure. The Ge concentration is varied during the film deposition process to provide a position-dependent SiGe alloy film for profile and device optimization of the SiGe HBT base region . The epitaxial base region forms a single-crystal SiGe intrinsic base and an amorphous poly-SiGe extrinsic base region. A window is formed over the single-crystal intrinsic SiGe base region to form the *n*-type polysilicon emitter. After hot process, n^+ dopants of the emitter diffuse into the SiGe intrinsic base region. Interconnection to the emitter, base and collector is defined by a tungsten (W) local interconnect. Interlevel dielectrics, tungsten contacts, and aluminum interconnects are formed with the reactive ion etching (RIE) and chemical mechanical polishing (CMP) processes used in base CMOS technology.

Heterojunction base–emitter design, bandgap engineering and technology scaling will each play a key role in the ability to achieve faster devices for the wired and wireless markets. As these structures are scaled, the sensitivity of these devices from EOS, ESD and EMI events. Figure 11.2 shows a cross-sectional drawing of a silicon–germanium heterojunction bipolar transistor.

In this structure, the deep trench structure borders the collector region. The collector region contains the n^{++} sub-collector, the lightly doped region of the collector and the n^+ pedestal implant. The shallow trench isolation is formed on the deep trench sidewall and serves as a region to define the base electrical contact. The collector contact is not shown in the drawing. The SiGe base region is formed over the single-crystal silicon and

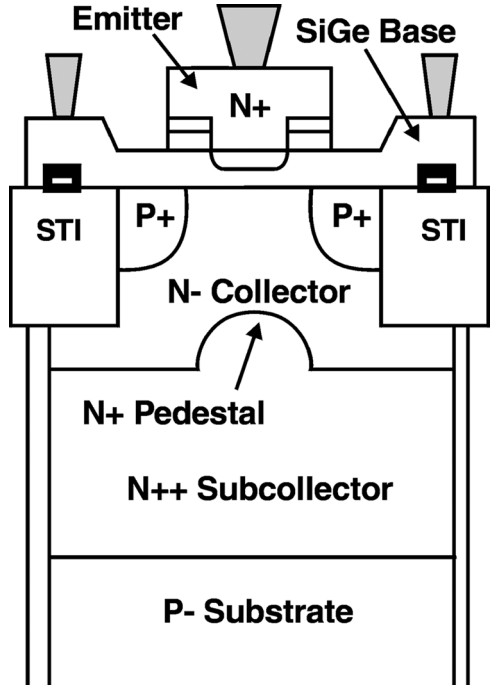

Figure 11.2 SiGe HBT device

the shallow trench isolation regions. Contacts to the base region are made over the shallow trench isolation regions to the poly-silicon–germanium film region. Dopants from the base diffuse into the single crystal region forming the base–collector junction region. The emitter is formed in the single-crystal silicon–germanium region by diffusion.

Epitaxial base profile optimization of the SiGe HBT device provides degrees of freedom for device design. These design freedoms can also translate into the ability to design a heterojunction transistor with electrical characteristics suitable for both functional optimization and high-current pulse operation (e.g. ESD operation, overshoot/undershoot). Today, SiGe HBT devices use a spatially dependent Ge concentration to modulate the bandgap, which provides a higher collector current density, base transit time reduction, low intrinsic base resistance, and an increased Early voltage at cutoff frequencies. Silicon–germanium HBT devices can be formed by varying the germanium concentration in the Si_xGe_{1-x} compound. Figure 11.3 shows the SiGe bandgap highlighting the band modification from the germanium profile. Rectangular, triangular and trapezoidal profiles are used to optimize the transistor transit time.

Figure 11.4 shows an expanded view of the emitter–base region. Over the single-crystal silicon, a pseudomorphic single-crystal SiGe film is formed. A facet occurs at the amorphous SiGe-to-SiGe single-crystal interface. The $TiSi_x$ salicide is formed on the extrinsic base region from the contact to the edge of the emitter polysilicon structure. A p^+ dopant is implanted into the extrinsic base region to reduce the base series resistance. A 'self-aligned' (SA) SiGe HBT device is a transistor where this extrinsic base implant is aligned to the edge of the emitter window opening using a disposable spacer forming a fixed extrinsic base-to-emitter space. A 'Non-self-aligned' (NSA) SiGe HBT device is a transistor where the extrinsic base implant is aligned to the edge of the polysilicon emitter shape.

Two key figures of merit for a SiGe HBT device are the unit current gain cutoff frequency f_T, and unity power gain cutoff frequency f_{max}. The unit current gain cutoff frequency f_T is inversely related to the emitter-to-collector transit time τ_{EC}.

$$\frac{1}{2\pi f_T} = \tau_{EC} = \tau_E + \tau_B + \tau_{CSL} + \tau_C$$

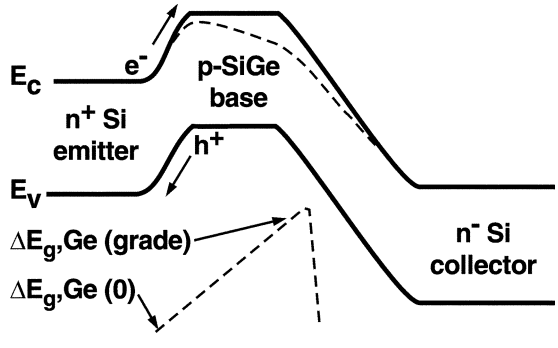

Figure 11.3 SiGe HBT band diagram

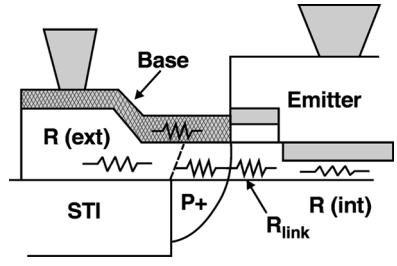

Figure 11.4 SiGe HBT emitter–base structure

where

$$\tau_E = \frac{C_{eb} + C_{bc}}{g_m}$$

$$\tau_B = \frac{W_B^2}{KD_B}$$

$$\tau_{CSL} = r_C(C_{c-sx} + C_{bc})$$

$$\tau_C = \frac{x_C}{\upsilon_{sat}L}$$

The emitter transit time is a function of the emitter–base capacitance, base–collector capacitance, and transconductance. The base transit time is a function of the base width, and diffusion coefficient and base grading factor K. The collector terms consists of the collector capacitance, collector resistance and velocity saturation term.

Minimizing the collector–base capacitance C_{cb}, collector–substrate capacitance C_{csx}, and base resistance R_b improves f_{max}.

$$f_{max} = \sqrt{\frac{f_T}{8\pi R_b C_{bc}}}$$

The base resistance is defined as an intrinsic base resistance and an extrinsic base resistance. The intrinsic base resistance is associated with the area under the n^+ emitter diffusion. The extrinsic base resistance consists of all base resistance components from the intrinsic base to the base contact. The extrinsic base resistance comprises of the amorphous SiGe region from the contact to the STI edge. At the STI–silicon transition, the SiGe film transitions from amorphous to single crystal SiGe. The first term is the resistance of the poly-SiGe. The second term is the resistance of the SiGe film from the facet to the edge of the extrinsic base implant. The third component of the extrinsic base implant is the single-crystal SiGe base region between the emitter diffusion edge to the extrinsic base implant edge, known as the 'link' resistance.

$$R_{b,int} = \frac{1}{3}\frac{W_E}{L_E}\rho_B$$

$$R_{\mathrm{b,ext}} = R_{\mathrm{polySiGe}} + R_{\mathrm{SiGe}} + R_{\mathrm{Link}}$$

$$R_{\mathrm{Link}} = \frac{L_{\mathrm{Link}}}{L_{\mathrm{E}}}\rho_{\mathrm{Link}}$$

The first two regions of the external resistances are a function of the salicide location relative to the facet transition. The salicide extends from the contact to the emitter polysilicon edge.

Optimization of the emitter–base region requires addressing the cutoff frequencies f_{T} and f_{max}, base resistance R_{b}, emitter–base breakdown voltage $\mathrm{BV_{EBO}}$, and emitter-base leakage current $\mathrm{IL_{EBO}}$. For SA SiGe HBT devices, there is an optimum spacer thickness which optimizes the emitter–base region. As the spacer width decreases, the cutoff frequencies decrease due to an increase of the emitter–base and collector–base capacitances. Additionally, the $\mathrm{BV_{EBO}}$ decreases and the $\mathrm{IL_{EBO}}$ increases. As the spacer width decreases, it has been shown that the leakage current changes from a temperature-sensitive temperature-insensitive behaviour. This indicates transition from thermal leakage mechanisms (e.g. Shockley–Read–Hall, Frenkel–Poole, and avalanche) to tunneling mechanisms (e.g. trap-to-band and band-to-band tunneling). Current gain degradation effects are observable from current gain measurements, Gummel plots and I_{B}–V_{BE} plots.

11.1.2 Silicon–Germanium Device Physics

The physics of a silicon–germanium HBT device compared with the silicon bipolar junction transistor is different as a result of the bandgap engineering. Epitaxial base profile optimization of the SiGe heterojunction bipolar transistors provides degrees of freedom for transistor device design. Kroemer pointed out in 1982 that heterojunctions can control electrons and holes separately and independently provide design freedom which is not achievable in homojunction transistors [8,9]. These design freedoms can also translate into the ability to design a heterojunction transistor with electrical characteristics suitable for both functional optimization and high-current pulse operation (e.g. ESD operation, overshoot/undershoot). Kroemer noted that using wide-bandgap emitter structures was extremely advantageous for high base doping concentrations in the heterojunction bipolar transistor [8,9]. Today, SiGe heterojunction bipolar transistors use a spatially dependent germanium concentration to provide a higher collector current density, base transit time reduction, low intrinsic base resistance, and an increased Early voltage at cutoff frequencies.

With high base doping concentrations, the intrinsic temperature T_{i}, of a SiGe HBT device can be significantly higher than that of a Si BJT device, providing a device less prone to thermal instability. The intrinsic temperature T_{i} is the temperature at which the intrinsic carrier concentration exceeds the dopant concentration. Thermal runaway or second breakdown is the point (I,V) at which a semiconductor device is at the onset of a negative resistance state because of thermal feedback initiated by Joule self-heating. The thermal instability is typically irreversible and leads to either leakage or destruction of the semiconductor structure. From Joule heating, the intrinsic carrier concentration

increases with increasing temperature. It is well known that Si-BJT device thermal instability typically occurs in the low-doped base region, where the intrinsic temperature is low, due to the low base doping concentration. For SiGe devices, the position-dependent Ge concentration also plays a role. With a position-dependent Ge concentration, the intrinsic carrier concentration can be expressed as

$$n_i^2(x) = \gamma n_{io}^2 e^{\Delta E(x)/kT}$$

and

$$\Delta E_g(x) = \Delta E_{g,b}^{app} + \Delta E_{g,Ge}(\text{grade})(x/W_b kT) + \Delta E_{g,Ge}^{(0)}$$

where $\Delta E_g(x)$ is the position dependent bandgap; $\Delta E_{g,b}^{app}$ is the heavy doping-induced apparent bandgap narrowing; and $\Delta E_{g,Ge}$ (grade) is the difference between the band bending at the base–collector junction; $\Delta E_g(x = W_b)$, and the band bending at the base–emitter junction; ΔE_g $(x = 0)$. The effective density of states ratio of SiGe and Si is expressed as

$$\gamma = (N_C N_V)_{SiGe}/(N_C N_V)_{Si}$$

The position-dependent Ge concentration produces a position-dependent intrinsic temperature $T_i(x)$. Hence, the intrinsic temperature for a heterojunction bipolar transistor can be expressed as

$$T_i(x) = \left\{E_{g,Si} + \Delta E_{g,b}^{app} + \Delta E_{g,Ge}(\text{grade})(x/W_b) + \Delta E_{g,Ge}^{(0)}\right\}/$$

$$k \ln\left\{[N_b(x)/n_{io}]^2 (N_C N_V)_{Si}/(N_C N_V)_{SiGe}\right\}$$

From an ESD perspective, a bipolar transistor with a high bipolar current gain is also advantageous for discharging current in a peripheral circuit or ESD network. From the position-dependent intrinsic carrier concentration and the Kroemer–Moll–Ross relationship [8,9], the collector current for a graded Ge concentration can be expressed as

$$I_C = A_C\left(qD_{nb}n_{io}^2/W_b N_b\right)\left\{e^{\Delta E_{g,b}^{app}/kT} e^{V_{be}/kT} - 1\right\}$$

$$\left\{\gamma\nu(\Delta E_g(\text{grade})/kT)/\left(1 - e^{-\Delta E_{g,Ge}(\text{grade})/kT}\right)\right\}$$

with γ and ν are position-averaged.

From the Kroemer–Moll–Ross collector–current relationship, the ratio of the current gain of a SiGe HBT and Si BJT can be expressed as

$$\frac{\beta_{SiGe}}{\beta_{Si}} = \frac{\gamma\nu(\Delta E_g(\text{grade})/kT) e^{\Delta E_{g,Ge}^{(0)}/kT}}{\left(1 - e^{\Delta E_{g,Ge}(\text{grade})/kT}\right)}.$$

In a common emitter mode, the collector current can be expressed as a function of the avalanche multiplication factor M and the current gain term as

$$I_C = \frac{M\beta_{SiGe}I_B + MI_{CO}(1+\beta_{SiGe})}{1-\beta_{SiGe}(M-1)}$$

where in an open-base condition

$$I_C = \frac{MI_{CO}(1+\beta_{SiGe})}{1-\beta_{SiGe}(M-1)}$$

From this expression, it is clear that the collector current is a strong function of both the avalanche multiplication factor M, and bipolar current gain. The high bipolar current gain of SiGe HBT transistors allows for the ability to discharge significant current during ESD events. The avalanche multiplication term M also plays a role in the high-current conduction during ESD events.

With bandgap engineering, the SiGe transistor is designed so that the Ge concentration does not significantly penetrate into the Si collector region at the base–collector junction. Germanium has a lower high-electric-field effective ionization threshold energy compared with Si (Ge is approximately 1 eV compared with 3.6 eV in Si), leading to a higher impact ionization rate. Fortunately, since the penetration of the Ge concentration is a few tens of nanometers, there is not a significant accumulation of carrier energy leading to an enhancement of the avalanche multiplication term M at the base–collector junction.

For ESD phenomenan, it is important to have the bipolar transistor responsive on the time scale of the ESD model. The CDM model, the fastest of the ESD models, has a 250 ps rise time and an oscillatory waveform. The energy spectrum of a CDM event has a significant energy contained in the 1–5 GHz regime. The responsiveness of the SiGe HBT device is a function of the emitter, base and collector transit times, whereas the base transit time is typically the frequency-limiting term. SiGe HBT devices use a graded Ge profile to establish a built-in electric field which accelerates the electrons from the emitter to the collector, providing a base transit time significantly less than Si BJT devices.

From the double-integral expression of the base transit time for a heterojunction, the base transit time can be expressed as

$$\tau = \left(W_b^2/D_{nb}\right)\left(kT/\Delta E_{g,Ge}(\text{grade})\right)$$
$$\left[1-\left(kT/\Delta E_{g,Ge}(\text{grade})\right)\left(1-e^{-\Delta E_{g,Ge}(\text{grade})/kT}\right)\right]$$

This expression is comprised of the base transit time in a uniformly doped Si BJT, W_b^2/D_{nb}, multiplied by the terms associated with the base transit time enhancements due to the Ge bandgap modulation. The emitter transit time in a SiGe HBT device is proportional to the reciprocal of the bipolar current gain, so it has the advantage relative to the Si BJT for ESD phenomenon.

Breakdown voltages of SiGe HBT devices are important for both I/O and ESD applications. The collector-to-emitter breakdown voltage, BV_{CEO} , can be expressed as

$$BV_{CEO} = BV_{CBO}\left\{1 - \frac{(\beta_{SiGe}/\beta_{Si})}{\left[\frac{\beta_{SiGe}+1}{\beta_{Si}}\right]}\right\}$$

For SiGe HBT devices, the bipolar collector current gain improvement shows itself as a low collector-to-emitter breakdown voltage. The collector-to-emitter breakdown voltage, $BV_{CEO} - \beta$ tradeoff can be advantageous for both performance and ESD protection by modulating the collector profile with a 'pedestal implant.' Using an optional pedestal implant, both a high-frequency f_T and a high-breakdown voltage SiGe *npn* can be built in a single process for I/O and ESD advantages.

11.2 SILICON–GERMANIUM ESD MEASUREMENTS

The sensitivity of silicon–germanium heterojunction bipolar transistors ESD needs to be evaluated in different bipolar configurations. Two SiGe HBT device measurements of interest are collector-to-emitter and emitter-to-base measurements.

In a collector-emitter mode, the emitter is grounded and an ESD pulse is applied to the collector region. As the collector-to-emitter voltage increases, both the collector–base junction voltage and the maximum electric field in the junction increases, leading to the onset of avalanche multiplication. In SiGe HBT devices, the base doping concentration is significantly higher than for the Si BJT devices. With the high-doped base region, the depletion region extends into the lower-doped collector region in both the pedestal and non-pedestal SiGe HBT devices. Avalanche multiplication occurs primarily in the Si collector region, where the holes are accelerated through the silicon collector into the SiGe base region and the electrons are accelerated into the collector. As the base region potential rises, the emitter–base junction becomes forward-biased, leading to bipolar snapback of the SiGe HBT device.

As an example measurement configuration, a transmission line pulse (TLP) system is shown (Figure 11.5).

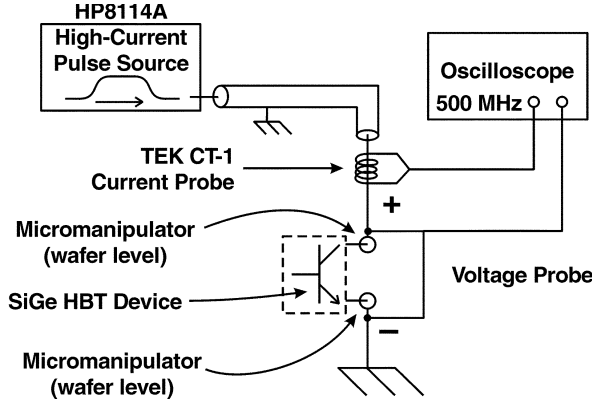

Figure 11.5 High-current transmission line pulse (TLP) test system

A high-current pulse-testing I–V measurement of a SiGe HBT with a 0.32-μm-wide by 20-μm-long emitter structure in a floating base common emitter mode is shown in Figure 11.6. For this measurement, the pulse has 50-ns pulse width with 3-ns rise- and fall-times. With single probes, no capacitive load exists on the base electrode.

As the TLP pulse current increases, the collector-to-emitter voltage increases across the SiGe HBT device until the first trigger voltage is reached. As the collector voltage increases, avalanche multiplication increases until avalanche breakdown occurs. From this I–V characteristic, the first trigger point of avalanche (I_{t_1}, V_{t_1}) is the onset of snapback. With increasing collector-to-emitter voltage, the current through the device increases, rising out of the low-voltage/high-current snapback state. The inverse slope of the I–V trace after the snapback point is the dynamic on-resistance of the SiGe HBT device. As the voltage increases, the current increases linearly with the increased pulse current. At this point, the voltage increases beyond the first trigger voltage, V_{t_1}. Beyond the first trigger point, the current increases until a soft transition occurs. Then, a change in the dynamic resistance occurs: the dynamic on-resistance is reduced, followed by an increasing current without a negative resistance transition or thermal instability. As the current pulse increases, the onset of thermal instability eventually occurs, where the SiGe HBT enters a negative resistance regime, and falls to a high-current/low-voltage state. The low-voltage/high-current state of the SiGe HBT is at a voltage value near or below the snapback voltage state. Typically, for most of the measurements taken in different configurations, the onset of thermal instability or the low-voltage/current state is the point of increase in leakage current and is believed to be the point of device failure.

Thermal instability and power-to-failure in high-current pulse testing is a function of the applied pulse width as is evident from the Wunsch–Bell and Dwyer models. As shown in the Wunsch–Bell model, the critical current-to-failure and power-to-failure of a semiconductor device increases with decreasing pulse width. Figure 11.7 shows TLP I–V characteristics of a $0.32 \times 20\,\mu m$ SiGe HBT as a function of pulse width. For pulse widths significantly shorter than the thermal diffusion time, the power to failure of a SiGe HBT will be inversely proportional to the pulse width and follow an adiabatic dependence. For pulse widths on the order of the thermal diffusion time, the critical

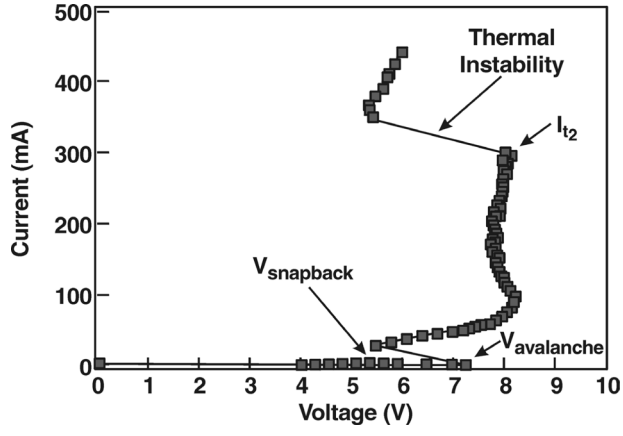

Figure 11.6 TLP I–V characteristic of SiGe HBT device (pulse width $\tau = 50\,ns$)

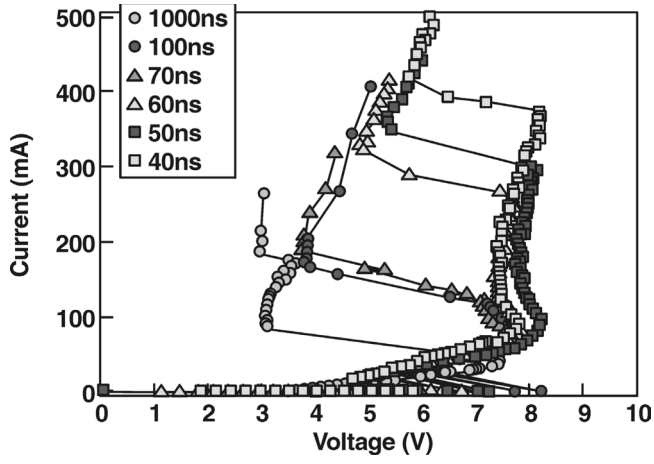

Figure 11.7 TLP I–V characteristic of SiGe HBT device ($\tau = 40$–1000 ns)

current to failure will have a weaker dependence on pulse width due to the thermal diffusion of heat from the region of self-heating. As the pulse width increases above 50 ns, our measurements show that the onset of thermal instability decreases monotonically with increasing pulse width. The post-thermal instability low-voltage/high-current I–V point also monotonically decreases in both current and voltage for the observed pulse width range from 40 to 1000 ns. An additional observation is that, for very short pulse widths, a high-current state exists for the SiGe HBT prior to onset of thermal instability whereas this is less evident in the longer pulse widths. For the dynamic on-resistance, the shorter pulse widths have a lower resistance than the longer pulse I–V measurements as the pulse width increases from 40 to 1000 ns.

From these results, the Wunsch–Bell curve can be plotted by mapping the point of the current to failure as a function of the pulse width. Figure 11.8 shows a linear logarithmic plot of the onset of thermal instability J_{crit}, as a function of the pulse width for the DT-defined high-performance $0.32 \times 20\,\mu$m emitter SiGe HBT structure for the

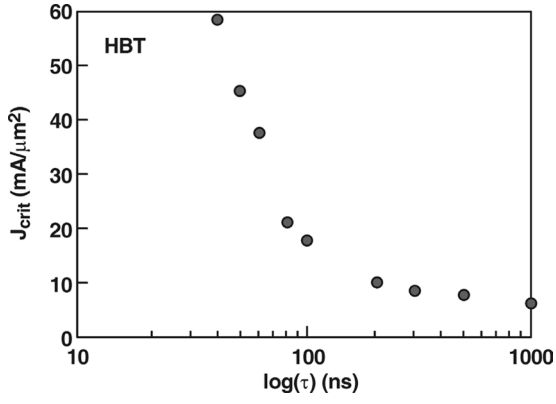

Figure 11.8 SiGe HBT critical current density J_{crit} as a function of pulse width τ

nine different pulse widths. The results show that J_{crit} decreases as the pulse width increases. The decrease in the onset of thermal instability and power to failure is consistent with work of Wunsch–Bell, and other established thermal physics failure models.

11.2.1 SiGe – Silicon Comparison

Silicon germanium transistors, compared with Silicon homojunction transitors have higher bipolar current gain characteristics, lower base transit time, lower emitter transit time, and lower intrinsic base resistance which are advantageous for ESD protection networks. An experimental batch with two epitaxial base Si homojunction splits and the standard SiGe HBT device was constructed with identical masks in common structures. Figure 11.9 shows a TLP *I–V* characteristic of one of the Si homojunction BJT splits in the identical $0.32 \times 20\,\mu m$ emitter structure for a 50 ns pulse width compared with the SiGe HBT measurement. The SiGe HBT device remains in a high-current/high-voltage state and does not undergo the transition into thermal instability, as is evident in the Si BJT device.

Further examination of this effect is provided by varying the pulse width from 40 to 1000 ns. Figure 11.10 shows the TLP *I–V* characteristics of the Si BJT homojunction. The response of the epitaxial base Si BJT homojunction *I–V* characteristic is similar to the SiGe HBT, with one key qualitative distinction. It is evident that the Si homojunction BJT undergoes an earlier transition into thermal instability for shorter pulse widths.

Figure 11.11 shows a Wunsch–Bell J_{crit} – pulse-width plot comparing the SiGe HBT and Si BJT device for a $0.32 \times 20\,\mu m$ transistor structure. For short pulse widths, the SiGe HBT has a higher J_{crit} than the Si BJT. For the longer pulse widths, the critical current density SiGe HBT and Si BJT converge. Wunsch–Bell current to failure plots also demonstrate that the SiGe HBT ESD results are 30% superior to the Si BJTs at 100 ns. At shorter times, the difference is 200%. This comparison was also completed for other Si BJT experimental splits where, in all cases, the SiGe HBT showed superior results at shorter time constants.

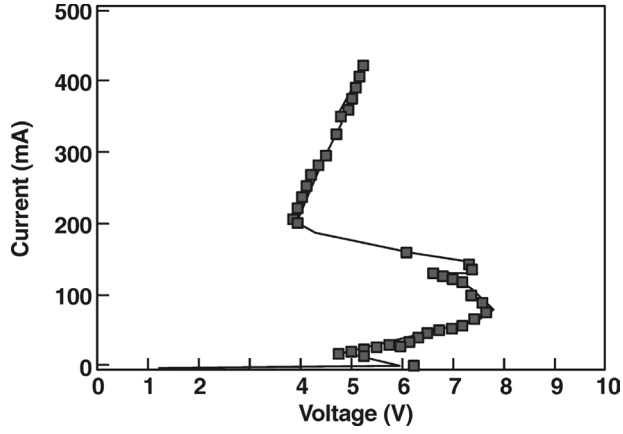

Figure 11.9 TLP *I–V* characteristic of silicon homojunction BJT ($\tau = 50$ ns).

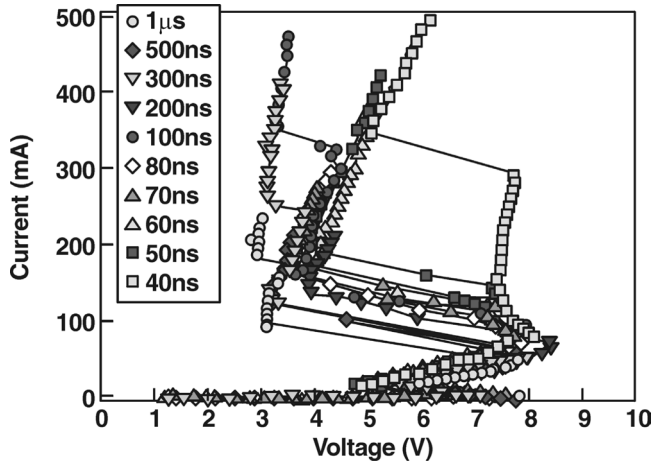

Figure 11.10 TLP I–V characteristic of silicon homojunction BJT ($\tau = 40$ ns to 1 μs)

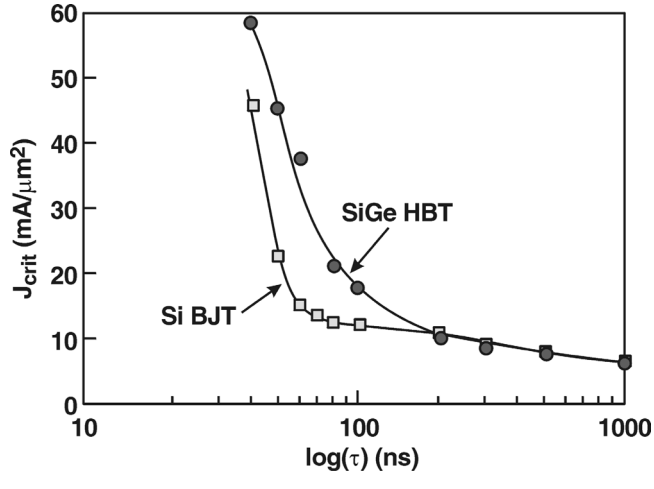

Figure 11.11 Wunsch–Bell J_{crit}–τ plot for a SiGe HBT and Si BJT device

The comparison of SiGe HBT and Si homojunction BJT low- and high-current characterization parameters is of considerable interest to circuit designers and device engineers who must optimize and trade off functional and ESD characteristics. In Table 11.1 the electrical parameters and ESD-related parameters are shown. Key parameters are the current gain β, the Early voltage V_A, breakdown voltages and second breakdown values. Two key metrics of interest are the current gain–Early voltage product βV_A, and a new hybrid metric of the current gain-second breakdown βI_{t_2} metric. The current gain–Early voltage metric provides a measure of the current drive and output impedance of a SiGe HBT. The current gain second breakdown metric is a measure of the functional current-drive capability and the ESD robustness of the structure.

Table 11.1 Electrical parameters of SiGe HBT and Si BJT devices

Parameter	Si BJT (a)	Si BJT (b)	SiGe HBT
β	56.1	40.83	144
V_A (V)	41.63	53.12	103
I_{coll} at 0.72 V (μA)	0.82	0.62	2.15
BVEBO (V)	3.67	3.14	3.86
BVCEO (V)	3.63	3.78	3.2
BVCES (V)	11.16	11.1	11.26
I_{t_2} (A)	0.08	0.08	0.12
P_f (W)	0.64	0.71	0.845
$\beta\ V_A$	2335	2169	14832
βI_{t_2}	4.48	3.38	16.58

For cases where the SiGe-to-Si β_{SIGE}/β_{SI} ratio is 2.57 and 3.53, the SiGe HBT transistors provide both a higher βV_A and βI_{t_2} compared with both epitaxial base Si homojunction and *npn* BJT devices. The $(\beta I_{t_2})_{SiGe}/(\beta I_{t_2})_{Si}$ ratio can be expressed as

$$\frac{(\beta I_{t_2})_{SiGe}}{(\beta I_{t_2})_{Si}} = \eta\left(\Delta E_{g,Ge}(\text{graded})/kT\right)e^{\Delta E_{g,Ge}^{(0)}/kT}$$

$$\left(1 - \exp^{-\Delta E_{g,Ge}(\text{graded})/kT}\right)\{I_{t_2 SiGe}/I_{t_2 Si}\}$$

From the Moll–Ross–Kroemer current relationships, the SiGe-to-Si β advantage is consistent with correlation of improved I_{t_2} with current gain β [8,9].

A key advantage of the bandgap engineering for the SiGe base region allows for design optimization of the intrinsic base resistance R_{bi} and cutoff frequency, f_T. The intrinsic base resistance for a SiGe HBT device can be expressed as

$$R_{bi(SiGe)}/R_{bi(Si)} = (\beta_{SiGe}/\beta_{Si})\left\{e^{-E_g^{(0)}/kT}\left(kT/\Delta E_g(\text{grade})\right)\right\}$$

A lower zero bias intrinsic base resistance can be achieved for a given frequency f_T, allowing for higher doping concentration in the intrinsic and extrinsic base regions. By providing a higher doping concentration, a higher intrinsic temperature, $T_i(x)$, can be obtained, improving the second breakdown characteristics and providing a thermally stable base region. This is a significant advantage for ESD robustness of SiGe HBT devices. Additionally, the lower intrinsic base resistance reduces the impact of voltage distribution drops in a SiGe *npn* stripe and lower resistance in a diode-configured base–collector *npn* transistor compared with a Si BJT so that they can be more suitable as ESD protection network elements. Successful usage of a SiGe HBT device in diode-like operation is also important to provide a low extrinsic base resistance and low collector resistance. The low extrinsic base resistance is minimized by design and device scaling whereas the low collector resistance is achieved with high sub-collector doping concentrations and reach-through implants.

11.2.2 SiGe Electrothermal Simulation: Collector–Emitter

Preliminary electrothermal simulation of the SiGe HBT device during an HBM ESD event was completed using a two-dimensional finite element cross-section of the SiGe HBT device. A full SiGe semiconductor process was defined, using the TSUPREME semiconductor process tool. Finite element electrothermal simulation was completed, using the FIELDAY III device simulator. The FIELDAY formulation is based on the three moments of the Boltzmann transport equations, conservation of charge equations, and lattice energy conservation relationships. The constitutive equations for electron- and hole-current densities account for current flow caused by thermal gradients. The constitutive equation for lattice energy flux represents the Fourier law of heat conduction. The lattice energy consists of the Joule heating and energy exchange with carriers and the lattice.

To simulate an HBM pulse, the Van Roozendaal double-exponential current waveform is used in the FIELDAY simulator. Figure 11.12 shows the temperature field in the SiGe HBT device where the cross-section was sliced through the center of the emitter structure, using a reflecting thermal boundary condition ($dT/dy = 0$) with thermal contacts and proper thermal boundary conditions established at other interfaces. In the study, the simulation was completed with the test in common emitter configuration. The emitter window is in the upper right-hand corner of the device. In the common emitter mode, the peak temperature can be observed in the collector region. In SiGe HBT devices, the base doping concentration is significantly higher than the collector region. As a result, the Joule heating in the collector region plays a significant role in the transient temperature field $T(x,y,t)$ in the SiGe HBT device. The darkest contour is the region of peak heating the collector–emitter region.

11.2.3 SiGe Transmitter–Emitter–Base

Emitter–base design (Figure 11.13) has a significant influence on both the device performance and ESD sensitivity of silicon–germanium heterojunction bipolar transistors [49,50]. The evaluation of process variations and device design spacings on ESD robustness is a function of the salicide location, emitter–base spacing, and collector

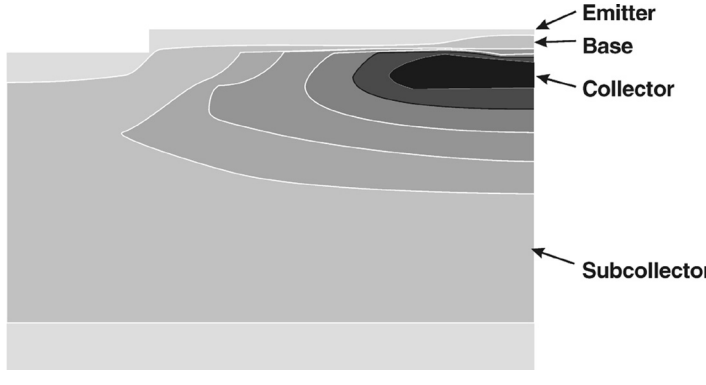

Figure 11.12 Transient electrothermal device simulation of SiGe HBT (HBM pulse)

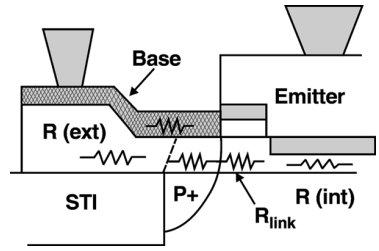

Figure 11.13 SiGe HBT emitter–base structure

opening. For high-speed RF applications, it is important to improve the ESD protection of RF receiver networks. A typical example of a RF circuit where RF (in) is connected to the base of a SiGe HBT device and RF (out) is connected to the collector. A key sensitivity of RF receivers is ESD failure of the emitter–base junction in both forward- and reverse-bias stress modes. The ESD robustness of the emitter–base junction is a function of the spacing, salicide, resistance and design of the emitter base structure.

Over the single-crystal silicon, a pseudomorphic single-crystal SiGe film is formed. A facet occurs at the amorphous SiGe-to-SiGe single-crystal interface. The $TiSi_x$ salicide is formed on the extrinsic base region from the contact to the edge of the emitter polysilicon structure. A p^+ dopant is implanted into the extrinsic base region to reduce the base series resistance. A 'self-aligned' (SA) SiGe HBT device is a transistor where this extrinsic base implant is aligned to the edge of the emitter window opening using a disposable spacer forming a fixed extrinsic base-to-emitter space. A 'non-self-aligned' (NSA) SiGe HBT device is a transistor where the extrinsic base implant is aligned to the edge of the polysilicon emitter shape.

The emitter transit time is a function of the emitter–base capacitance, base–collector capacitance, and transconductance. The base transit time is a function of the base width, and diffusion coefficient and base grading factor K. The collector terms consists of the collector capacitance, collector resistance and velocity saturation term. The base resistance is defined as an intrinsic base resistance and an extrinsic base resistance. The intrinsic base resistance is associated with the area under the n^+ emitter diffusion. The extrinsic base resistance consists of all base resistance components from the intrinsic base to the base contact. The extrinsic base resistance comprises the amorphous SiGe region from the contact to the STI edge. At the STI–silicon transition, the SiGe film transitions from amorphous to single-crystal SiGe. The first term is the resistance of the poly-SiGe. The second term is the resistance of the SiGe film from the facet to the edge of the extrinsic base implant. The third component of the extrinsic base implant is the single-crystal SiGe base region between the emitter diffusion edge to the extrinsic base implant edge, known as the 'link' resistance.

$$R_{\mathrm{b,int}} = \frac{1}{3}\frac{W_{\mathrm{E}}}{L_{\mathrm{E}}}\rho_B$$

$$R_{\mathrm{b,ext}} = R_{\mathrm{polySiGe}} + R_{\mathrm{SiGe}} + R_{\mathrm{link}}$$

$$R_{\mathrm{link}} = \frac{L_{\mathrm{link}}}{L_{\mathrm{E}}}\rho_{\mathrm{link}}$$

The first two regions of the external resistances are a function of the salicide location relative to the facet transition. The salicide extends from the contact to the emitter polysilicon edge [49,50].

Optimization of the emitter–base region requires addressing the cutoff frequencies f_T and f_{max}, base resistance R_b, emitter–base breakdown voltage BV_{EBO}, and emitter–base leakage current IL_{EBO}. For SA SiGe HBT devices, there is an optimum spacer thickness which optimizes the emitter–base region. As the spacer width decreases, the cutoff frequencies decrease, due to an increase of the emitter–base and collector–base capacitances. Additionally, the BV_{EBO} decreases and the IL_{EBO} increases.

For the self-aligned (SA) SiGe HBT device, as the emitter polysilicon width increases, the salicide is moved away from the emitter diffusion area while the extrinsic base implant is fixed. As the collector width increases, the length of the single-crystal SiGe region of the extrinsic base resistance increases yet the link resistance and the intrinsic base region remains fixed. Figure 11.14 shows HBM ESD results of a $0.44 \times 10\,\mu m$ SA SiGe HBT device where the salicide-to-emitter spacing was decreased below nominal spacing while the extrinsic base implant was kept fixed (in this case, the extrinsic base implant is self-aligned to the emitter window). For the SA SiGe HBT device, it was found that the base resistance was modulated by removal of the salicide film, leading to an increase in the positive stress base–emitter ESD results. With removal of the salicide film, current must flow through the polysilicon SiGe film instead of through the salicide film itself. In a self-aligned SiGe HBT, the increase in the size of the emitter polysilicon increases the base series resistance. The link resistance remains constant while the other two extrinsic terms are modified dependent on whether the salicide extends to the single-crystal Si region or over the isolation (e.g. the facet transition region). As the silicide is pushed out beyond the facet, the base resistance will increase significantly, exposing the amorphous SiGe film.

From the unit current gain–collector current ($f_T - I_C$) and unit power gain–collector current ($f_{max} - I_C$) RF measurements, the peak $f_{max} - I_C$ characteristic shows that the peak f_{max} decreases with increasing spacing of the salicide from the silicon emitter. This effect can be demonstrated by increasing the polysilicon emitter width or by using

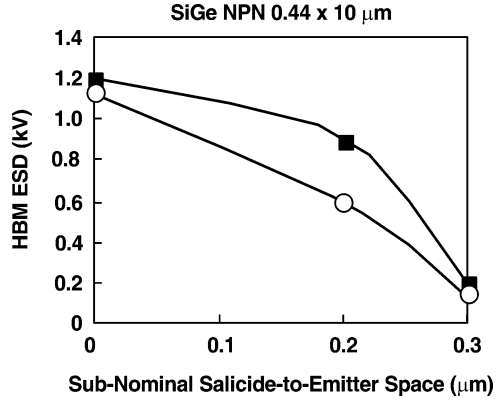

Figure 11.14 SiGe HBT emitter–base HBM ESD robustness as a function of salicide-to-emitter spacing

a salicide block mask. Using a salicide block mask level, RF measurements show an insignificant change in the peak f_T when the salicide is removed from within 0.4 μm of the emitter polysilicon edge. With increasing the base series resistance, the peak f_{max} decreases from 45 to 42 GHz. This can be expressed as [49,50]

$$\frac{\partial f_{max}}{\partial R_b} = -\frac{f_{max}}{2R_b}$$

For a 'non-self aligned' transistor, the space between the extrinsic base implant and the emitter can be varied. As discussed previously, this has a significant influence on the emitter–base breakdown voltage BV_{EBO} and the emitter–base leakage current IL_{EBO}. In the NSA SiGe HBT, the extrinsic base implant and the silicide moves with the emitter polysilicon width. In our test structures, the space between the emitter polysilicon shape and the emitter window was varied.

In the NSA SiGe HBT device, the link resistance increases as the emitter polysilicon width increases and the salicide is moved farther from the emitter. During an ESD event, as the reverse-bias is applied, the emitter–base depletion width increases. Since the emitter is heavily doped, the depletion region extends laterally through the *p*-doped SiGe link resistance region. With increased reverse bias, the depletion region extends into the p^+ doped SiGe extrinsic base implant. Hence the NSA SiGe HBT device ESD robustness will be a strong function of the extrinsic base to emitter spacing.

For negative HBM stress across the emitter–base junction, experimental results show that as the emitter–base spacing is decreased, there is an decrease in the ESD robustness of the emitter–base junction (Figure 11.15). For HBM ESD stressing, the ESD failure criteria is a shift in the *I–V* characteristic. The failure is associated with an increase in the emitter–base leakage current, IL_{EBO}.

For positive-mode HBM pulses, the ESD results degrade as the link resistance is increased. Figure 11.16 shows the HBM ESD results for the forward-bias as a function of the extrinsic base encroachment. At very small extrinsic base–emitter spacing,

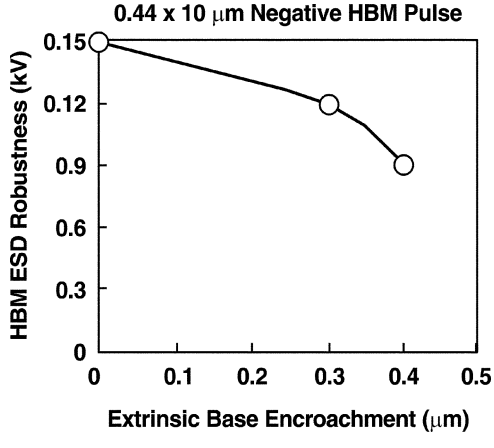

Figure 11.15 SiGe HBT emitter–base HBM ESD robustness as a function of the extrinsic base implant encroachment for a negative HBM pulse

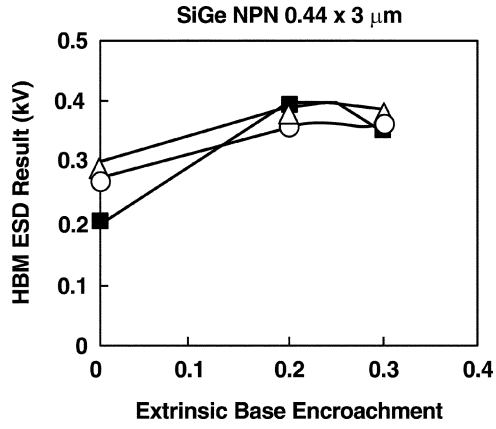

Figure 11.16 NSA SiGe HBT device HBM measurement in a forward-bias voltage stress as a function of extrinsic base encroachment

enhanced leakage current is evident, impacting the low current gain parameters. In Figure 11.16, it can be seen that as the link resistance increases, the ESD robustness of the base–emitter junction decreases [49,50].

From TLP measurements, it can be seen that the failure point and failure mechanism are associated with a significant increase in the current at a breakdown voltage. The TLP measurement of a NSA SiGe HBT device under reverse-bias emitter–base stress is shown in Figure 11.17. As the TLP applied current pulse increases, the emitter–base voltage increases. The increase in the emitter–base current is also associated with a small increase in the leakage current. The leakage current saturates at some value prior to the breakdown of the emitter–base junction.

At some value, the emitter–base junction undergoes a sudden increase in the reverse-bias emitter–base current as well as a significant jump in the leakage current [50]. A key result is that both leakage current and emitter–base current increase after approximately 4–5 V across the emitter base junction. This is beyond the safe operating area (SOA) of

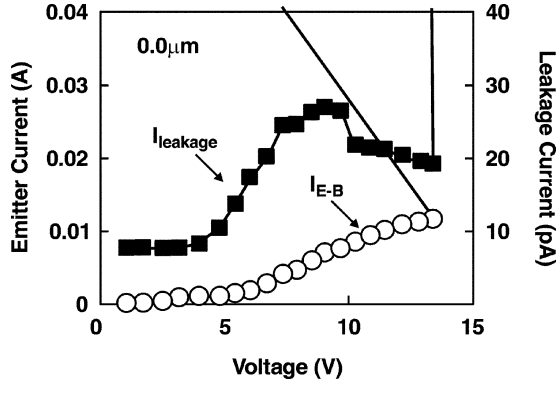

Figure 11.17 TLP measurement of NSA SiGe HBT emitter–base for large emitter diffusion-to-extrinsic base implant spacing

the SiGe HBT device. After 10 V, the leakage current saturates prior to final leakage increase. From the results in Figure 11.9, the emitter–base leakage current IL_{EBO} increases with the emitter–base current after the voltage exceeds the SOA voltage followed by a saturation phenomenon. A significant jump in leakage occurs until the avalanche breakdown voltage point. In Figure 11.18, the extrinsic base encroachment is increased by 0.3 μm. In this case, the SiGe HBT NSA device shows an increase in the emitter base current with the TLP pulse increase. Leakage current IL_{EBO} again increases after 4 V, followed by a saturation at 6 V. The IL_{EBO} has a sudden increase at the emitter–base current avalanche point.

In the NSA SiGe HBT, as the extrinsic base implant to emitter window edge is reduced, the breakdown voltage will reduce. In Figure 11.19, TLP measurements of the minimum extrinsic base implant to emitter window space NSA SiGe HBT device are shown

In Figure 11.19, for the 0.4 μm encroachment case, the failure voltage reduces to 5.05 V. In Table 11.2, the emitter–base encroachment and corresponding TLP failure voltage is shown with the leakage value prior to the failure of the structure.

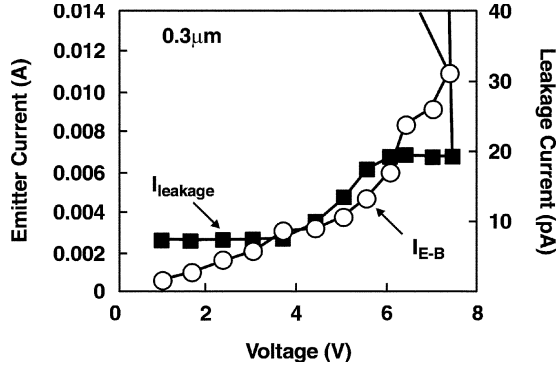

Figure 11.18 NSA SiGe HBT TLP measurement

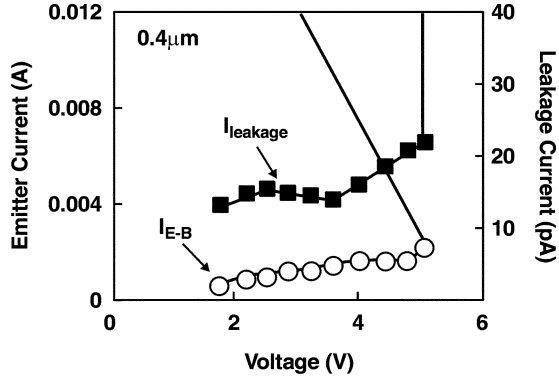

Figure 11.19 TLP measurement of NSA SiGe HBT emitter–base for minimum emitter-to-extrinsic base spacing

Table 11.2 TLP measurement summary of the NSA SiGe HBT emitter–base for minimum emitter-to-extrinsic base spacing

Emitter base encroachment (μm)	TLP leakage (pA)	TLP failure voltage (V)
0.00	19.3	13.487
0.20	12	10.453
0.30	19.83	7.41
0.40	18.0	5.046

Figure 11.20 shows the TLP failure voltage has a monotonically decreasing value with emitter–base encroachment.

A thermal model to explain the emitter–base failures would consist of a parallelepiped volume whose dimensions are the emitter length, the SiGe epitaxial thickness and the distance from the base contact to the emitter diffusion. The model consists of the amorphous poly-SiGe region, the SiGe single-crystal extrinsic base region, the low-doped SiGe *p*-region, the emitter n^+ diffusion and the SiGe epitaxial thickness. To analyze the reverse breakdown, a one-dimensional model for the $p^+/p^-/n^+$ with the appropriate link resistance lengths is suitable to understand the reverse voltage phenomenon. For the forward case, a one-dimensional Wunsch–Bell or three-dimensional Dwyer model with three thermal diffusion time constants associated with the SiGe epitaxial thickness, the emitter length and a physical length less than the base contact to emitter diffusion distance would be appropriate to understand the thermal physics in this structure; for a reverse case, a junction breakdown model which addresses the $p^+/p^-/n^+$ transitions.

A thermal transmission line model can also be used to explain the boundary conditions and self-heating in the polysilicon germanium film. The polysilicon germanium film is above the device surface and is on the shallow trench isolation region. Analogous to the silicon on insulator (SOI) model, the self-heating in the polysilicon germanium region can be represented as a thermal node, where there is an admittance expression

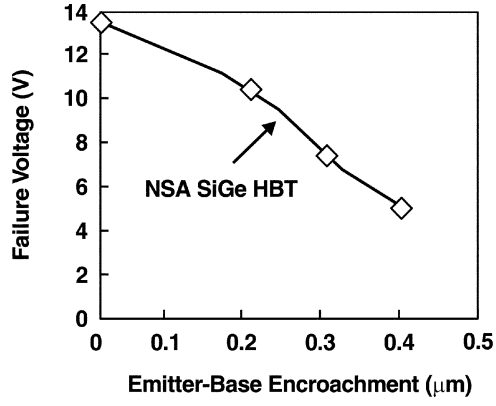

Figure 11.20 TLP failure voltage as a function of emitter–base encroachment in a NSA SiGe HBT device

consisting of the thermal capacitance and thermal resistance to the substrate region. The thermal capacitance and resistance of the shallow trench isolation will serve as the first admittance term. The boundary conditions can be represented as the base contact, and metal interconnects. From this model, the series solution can be obtained, and prediction of the peak temperature can be made.

11.3 SILICON–GERMANIUM–CARBON

Silicon–germanium alloy films have enabled the integration of advanced silicon technology with the heterojunction bipolar transistors devices [56–74]. In order to achieve higher speeds, new processes have emerged to combat the technology scaling limitations of the silicon–germanium HBT devices. One of the scaling limitations of the SiGe *npn* HBT device is the *npn* transistor base width. In order to reduce the transit time, the base width is required to scale to provide an increase in the unit current gain cutoff frequency f_T. Additionally, the base doping concentration must be increased to achieve a high unit power gain cutoff frequency f_{max}, to provide a low base sheet resistance as the base width is scaled. A limitation to achieve this performance is the use of boron in the base region. Boron diffusion will limit the ability to continually scale the SiGe HBT *npn* device base region. As a result, the silicon–germanium–carbon heterojunction bipolar transistor was introduced in order to address this issue, allowing the continued scaling of the SiGe HBT device to higher cutoff frequencies.

11.3.1 Silicon–Germanium–Carbon Device Physics

In order to achieve higher speeds, new processes have emerged to combat the technology scaling limitations of the silicon–germanium HBT devices. To extend the SiGe HBT device to higher frequencies, carbon is incorporated into its base region. A key issue in the technology scaling of the HBT device is the dimensional scaling of the base width of the transistor to achieve smaller base transit times. A challenge in pseudomorphic epitaxial base SiGe HBT devices is to scale the narrow as-grown boron profile within the SiGe alloy film during post-epitaxial processing. The challenge comes in achieving a low sheet resistance, by increasing doping concentration, and at the same time, limiting the boron diffusion process from thermal processes and transient enhanced Diffusion (TED) processes [56–74].

In a pseudomorphic $Si_{1-x}Ge_x$ film, the film is compressed in the growth plane and expanded in the growth direction, causing the boron atom diffusion to be nonrandom. Boron will diffuse more easily in the direction parallel to the silicon surface and diffuse more slowly in the film growth direction. The diffusion of Boron in the $Si_{1-x}Ge_x$ HBT alloy film can be explained in a simple diffusion model that includes the modified strain, the effect of trapping between B and Ge, the drift field due to the bandgap narrowing, and the spatially dependent intrinsic carrier concentration. The universal diffusion equation for a generic impurity present in a structure is

$$\frac{\partial C_B}{\partial t} = -\nabla \bullet J_B$$

where C_B is the boron concentration and J_B is the boron impurity flux, given by

$$J_B = -D_B\varphi\left(\frac{\partial C_B}{\partial x} + \frac{C_B}{Q}\frac{\partial N_T}{\partial x} + \frac{2C_B}{Q}\frac{n\partial \ln n_i}{\partial x}\right)$$

$$Q = \sqrt{(N_T^2 + 4n_i^2)} = n + p$$

$$\varphi = \frac{1 + \beta(p/n_i)}{1 + \beta}$$

and n and p are the electron and hole concentrations. The factor ϕ allows for the inclusion of the concentration dependence of both B and Ge fractions (strain) dependence of the effective diffusion coefficient. Additionally, the expression for ϕ contains the parameter β, and N_T and n_i are the net doping and intrinsic concentration, respectively. In the boron impurity flux equation, the first term is the conventional Fickian diffusion term, the second is a term for the built-in electric field from the dopant distribution, and the third term is associated with the built-in electric field initiated by the spatial variation of the $Si_{1-x}Ge_x$ bandgap. The diffusion of boron in a $Si_{1-x}Ge_x$ system is influenced by the different Ge fractions, the boron density gradient leading to a segregation process, and the electrical active B percentage (which is approximately half of a silicon-only film). Germanium acts as traps for the boron atoms which can be represented as a Ge–B clustering center for a boron clusterization process, which transfers B atoms into electrically inactive charge states.

$$B^- + Ge \rightarrow GeB^-$$

The diffusion of boron can be represented as

$$D_B = \varphi D_{BO}\exp(-E_A/kT)$$

At high Ge contents, the strain due to the presence of Ge influences the Boron diffusivity. The local strain, S, due to the Ge and B concentration is addressed through the modification of the activation energy

$$D_B = \varphi D_{BO}\exp(-E_A/kT)^*\exp(-fS/kT)$$

$$E_A = E_{A,mod} + fS$$

The diffusion of boron in the presence of carbon is significantly altered in the $Si_{1-x}Ge_x$ film. Carbon diffusion in regions with significant carbon concentrations can cause localized under-saturation of the silicon self-interstitials. This C-rich region results in the suppression of the boron diffusion process. The diffusion of C in silicon occurs as a result of a substitutional–interstitial mechanism. Silicon self-interstitials I react with immobile substitutional carbon C_S, to form mobile interstitial carbon atoms C_I, as represented in the reaction

$$C_S + I \Leftrightarrow C_I$$

The boron diffusivity is modified by the concentration of Si interstitials C_I, and the equilibrium concentration, $C*_I$.

$$D_B = \left(\frac{C_I}{C*_I}\right)\varphi D_{BO}\exp(-E_A/kT)$$

$$\varphi = \frac{1+\beta(p/n_i)}{1+\beta}$$

The boron activation energy in the equation above is a function of both the boron and the carbon. The activation is represented in the $Si_{1-x-y}Ge_xC_y$ system as $E_A = 3.62 + 0.4\,xy$, where x is the Ge and y is the C atom atomic fractional percentage [75]. Carbon, because the atom is small, relieves the strain in the $Si_{1-x}Ge_x$ film. From the relationship

$$D_B = \varphi D_{BO}\exp(-E_A/kT)^*\exp(-fS/kT)$$

the strain, S, can be represented as

$$S = |S_{Ge} - S_C|$$

$$S_{Ge} = \delta x$$

$$S_C = R\delta y$$

where δ is the lattice mismatch between Si and Ge, R is the strain compensation ratio, S_{Ge} and S_C are strain for Ge and C, respectively. The activation per unit strain f' is a function of the diffusivity of the strained and unstrained materials

$$f' = -\frac{kT}{S}\ln\left(\frac{\hat{D}}{D}\right)$$

11.4 SILICON–GERMANIUM–CARBON ESD MEASUREMENTS

11.4.1 SiGe Carbon Collector–Emitter Measurements

A SiGe system with carbon incorporation in the base region of a SiGe HBT device influences both the vertical and lateral transport of the boron in the intrinsic and extrinsic base region [75]. The reduction of the boron diffusion in the SiGe base regions allows for a narrower base width for a given base doping concentration. This provides a lower standard deviation of the base width and the base sheet resistance. The first case provides a higher unit current gain cutoff frequency f_T, and the second case allows for a higher unit power gain cutoff frequency f_{max}. Additionally, the retardation of the boron diffusion allows for the design reoptimization to increase the base doping concentration. The lowering of the base sheet resistance allows for a higher f_{max}. With the addition of carbon in the base region, the boron doping concentration can be increased to

provide a lower sheet resistance. For ESD robustness, there are two advantages. A higher base doping concentration will also improve the thermal stability of the SiGe HBT device as a result of the increased intrinsic temperature T_i, in the base region. Additionally, the higher boron concentration will also reduce the voltage drops in the base region leading to an improved current uniformity in the SiGe HBT during high current and ESD events.

In silicon bipolar transistors, statistical process variations lead to variations of the power to failure of transistor elements. Assuming a normal distribution, Pierce and Mason assumed the power to failure probability distribution function has a form

$$f_{P_f}(P_f) = \frac{1}{\sqrt{2\pi}S_p}\exp\left\{-1/2\left[\frac{P_f - \langle P_f\rangle}{S_p}\right]^2\right\}$$

where the power to failure P_f is the random variable and the mean power to failure is $< P_f >$, with standard deviation S_p. The standard deviation can be expressed as

$$S_p = \left[\frac{1}{N-1}\sum_{i=1}^{N}\{P_{f_i} - \langle P_f\rangle\}^2\right]^{1/2}$$

Pierce and Mason showed that the probability of failure occurs when the ESD pulse power exceeds the device power to failure. This can be expressed as a cumulative distribution function.

$$\Pr\{P_f > P_{f'}\} = \int_{P_{f'}}^{\infty} f_{P_f}(P_f, \langle P_C\rangle/\alpha, \langle P_C\rangle/\alpha\lambda)\,dP_f$$

In the Pierce–Mason model, the assumption in the analysis is that the parameter initiating the statistical variation in the failure sensitivity was the emitter junction depth. In our results, the power to failure variation will also be a function of the base width, doping distribution, sheet resistance and the thickness of the pseudomorphic $Si_{1-x}Ge_x$ film. Without carbon, the base width variation is dependent on the boron diffusion physics and implant statistical variations. The variations of the base width will be reflected in the collector–emitter breakdown voltage, and the base sheet resistance.

Figure 11.21 shows a typical common emitter TLP I–V measurements comparing a $0.44 \times 3.2\,\mu m$ SiGe and SiGeC HBT 45 GHz npn structure [75]. In this measurement, the pulse current is applied to the collector with the emitter grounded and the base floating. As the voltage increases in the collector region, avalanche multiplication occurs. In our results, current conduction is evident at the collector–emitter breakdown voltage BV_{CEO}. Snapback to a low current state is not evident after the breakdown voltage. This high-voltage/high-current regime is followed by a negative resistance state transition and ultimately an increase in the leakage current and device failure.

Figure 11.22 and 11.23 show the comparison of the voltage to failure of SiGeC and SiGe HBT devices. A key conclusion is the SiGeC ESD results clearly show a significant tighter ESD distribution in all experimental splits performed. Parametric data shows that the base resistance control is superior when carbon is incorporated into the base of

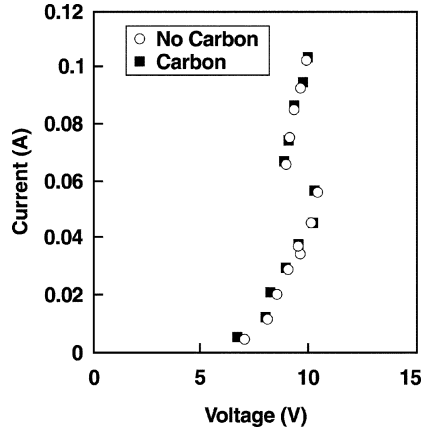

Figure 11.21 SiGe versus SiGeC HBT TLP characteristic (common emitter configuration)

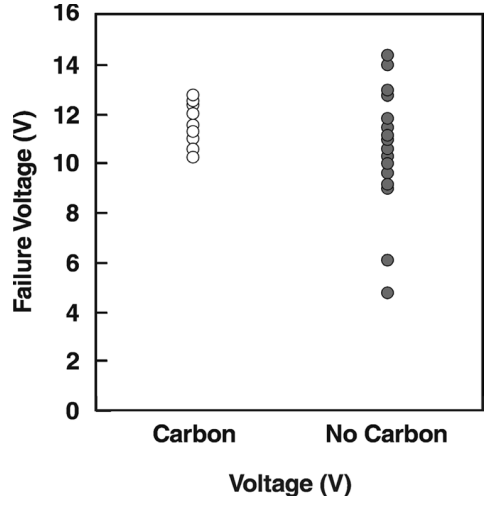

Figure 11.22 SiGe HBT voltage to failure with and without carbon incorporation

the SiGe *npn* HBT device, as anticipated because of improved control over the boron TED as well as improved Schuppen factor. This is evident from the d.c. electrical distribution of the base pinch resistance monitors. As anticipated, the SiGeC HBT provides improved ESD control due to improved base resistance distribution.

Figure 11.24 shows the comparison of the failure current histogram of the $Si_{1-x}Ge_x$ with and without carbon in the base region [75]. As was evident from the failure voltage histogram, the device with carbon is better controlled, providing an improved standard deviation.

Experimental results show that the relationship for the power of failure follows a normal distribution.

Experimental measurements were taken of the $Si_{1-x}Ge_x$ and $Si_{1-x-y}Ge_xC_y$ HBT devices as a function of the emitter width in a common emitter configuration.

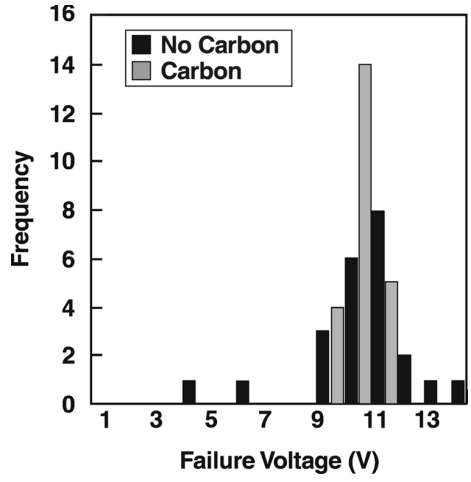

Figure 11.23 Histogram of SiGe HBT voltage to failure with and without carbon incorporation

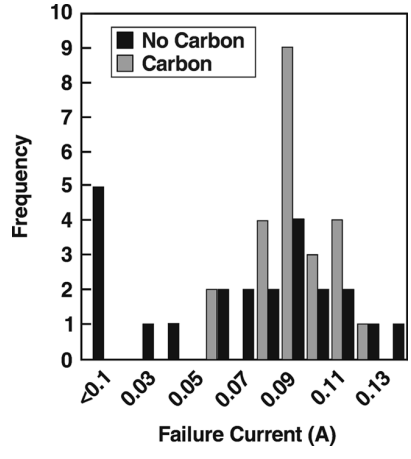

Figure 11.24 Histogram of SiGe HBT current to failure with and without carbon incorporation

Figures 11.25 and 11.26 show the TLP failure current and failure power as a function of the emitter width, respectively.

Our results show that for the $Si_{1-x}Ge_x$ and $Si_{1-x-y}Ge_xC_y$ HBT devices the ESD results increase linearly with increasing emitter width. At an emitter width of 0.35–0.45 μm, the TLP current to failure and power to failure begin to saturate.

11.4.2 SiGeC Device Degradation

During ESD stressing, gradual degradation is evident when the transistor is driven above the safe operation area (SOA). This was evident from the TLP measurements in the emitter–base mode of operation. The 'failure' of the transistor is dependent on the

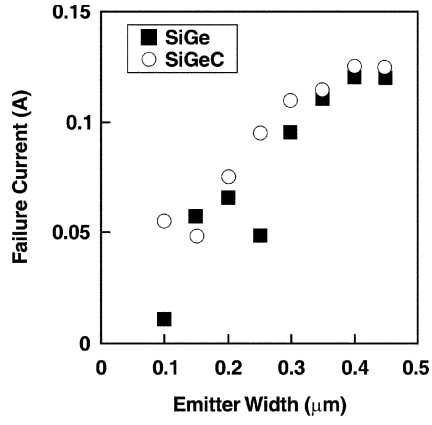

Figure 11.25 SiGe and SiGeC HBT TLP current to failure as a a function of HBT emitter width

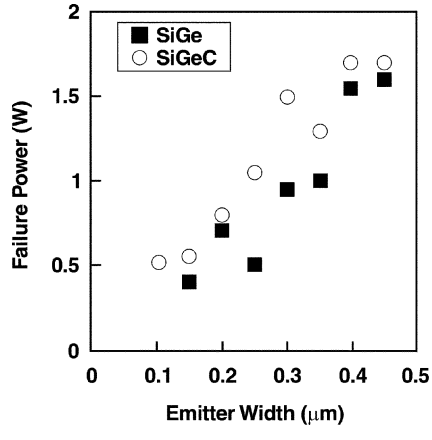

Figure 11.26 SiGe and SiGeC HBT TLP power to failure as a function of HBT emitter width

failure criteria. In RF devices, the failure criterion is a function of failure of the RF functional specifications, or transistor leakage.

In evaluation of the ESD robustness of $Si_{1-x-y}Ge_xC_y$ HBT device degradation, the ESD pulse can be varied in magnitude to observe the shift in the electrical parametrics. Electrical parameters can include analysis of the *S*-parameter, f_T–I_C, f_{max}–I_C, β degradation, Gummel plots and the d.c. parameter shift as a function of ESD pulses in forward and reverse polarity.

For evaluation of the SiGeC HBT device degradation, SiGeC HBT devices are to be placed in RF scattering parameter structures (e.g. *S*-parameter pad set with de-embedding test structures for RF characterization). A method to determine the degradation requires a test methodology including RF characterization both pre- and post-HBM wafer level stress

For the common emitter mode, a positive HBM pulse can be applied to the SiGe HBT after RF characterization. Our results showed no significant shift in the d.c. *I–V*

characteristic occurred until failure of the device. RF evaluation of f_T, and f_{max}, minimal degradation effects occurred prior to the device failure. Scattering parameters also failed to show significant degradation post-ESD testing. The results showed that no d.c. shift or RF degradation is evident in this device configuration prior to the complete device failure. At the failure point, the f_T and f_{max} characteristics show significant degradation where only a small peak was evident at high currents. The f_T–I_C and f_{max}–I_C shows significant decay in the lower current region prior to the peak at the failure point.

In the emitter–base configuration, d.c. and RF degradation is evident prior to the failure point and a relationship between the d.c. shift and RF degradation is evident in both the positive and negative pulse studies (Figure 11.27, Tables 11.3 and 11.4). During ESD testing, d.c. characterization showed a gradual degradation of the forward-bias *I–V* characteristic. By plotting the experimental results on a unit current gain cutoff frequency f_T against the collector current I_C, the change in the device characteristic can be determined. Evaluation of the $f_T - I_C$ plot shows as the d.c. shift occurs in the *I–V* characteristic, the peak current gain cutoff frequency decreases. The $f_T - I_C$ plot shows three effects. First, for current levels below the peak f_T, the current gain cutoff frequency is decreased with a larger slope in the df_T/dI_C term. Second, the peak f_T is decreased. Third, f_T beyond the peak f_T is weak variation with a convergence of the df_T/dI_C term.

Both positive and negative HBM pulses demonstrated RF f_T–I_C and f_{max}–I_C degradation and d.c. *I–V* forward characteristic shifts. From the Gummel plots, as the d.c.

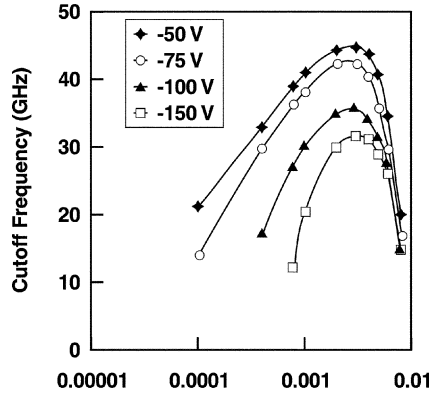

Figure 11.27 Unit current gain cutoff frequency versus collector current plot as a function of negative HBM pulse for SiGeC HBT emitter–base for different pulse magnitudes

Table 11.3 RF and d.c. Parameter degradation as a function of emitter–base negative HBM Stress

HBM pulse (V)	Forward voltage (V)	Peak cutoff frequency (GHz)
−50	0.72	46.49
−75	0.42	42.59
−100	0.38	34.78
−150	0.05	32.47

Table 11.4 RF and d.c. Parameter degradation as a function of positive HBM Stress

HBM pulse (V)	Forward voltage (V)	Peak cutoff frequency (GHz)
+50	0.78	46
+100	0.78	44
+200	0.78	44
+240	0.30	36
+250	0.18	31

degradation increases, the base current increases, and the base ideality factor (slope) decreases. The collector current does not shift significantly with the d.c. degradation. From the Gummel plots, the d.c. current gain β decreases with the increased ESD stress level. From this study, some significant conclusions can be made on the relationship of the d.c. shifts, the degradation mechanism, the RF parameter degradation and the condition of final HBM failure.

From the fact that unit current gain cutoff frequency f_T is inversely related to the emitter–collector transit time, τ_{EC}

$$\frac{1}{2\pi f_T} = \tau_{EC} = \tau_E + \tau_B + \tau_{CSL} + \tau_C$$

where

$$\tau_E = \frac{C_{eb} + C_{bc}}{g_m}$$

$$\tau_B = \frac{W_B^2}{KD_B}$$

$$\tau_{CSL} = r_C(C_c - sx + C_{bc})$$

$$\tau_C = \frac{x_C}{\nu_{sat} L}$$

From this expression, we can solve for the unity current gain cutoff frequency. Taking the derivative of the cutoff frequency equation with respect to the collector current,

$$\frac{1}{2\pi f_T} = \tau_{EC} = \tau_E + \tau_B + \tau_{CSL} + \tau_C$$

$$\frac{\partial}{\partial I_C}\left(\frac{1}{2\pi f_T}\right) = -\frac{1}{2\pi f_T^2}\frac{\partial f_T}{\partial I_C} = \frac{\partial}{\partial I_C}(\tau_{EC})$$

From this expression, only the emitter transit time transconductance term is a function of the collector current, hence this can be expressed as

$$\frac{\partial}{\partial I_C}\left(\frac{1}{2\pi f_T}\right) = -\frac{1}{2\pi f_T^2}\frac{\partial f_T}{\partial I_C} = \frac{\partial}{\partial I_C}(\tau_E) = \frac{\partial}{\partial I_C}\left(\frac{C_{eb} + C_{bc}}{g_m}\right)$$

where the change in the slope due to degradation can be expressed as

$$\frac{\partial f_T}{\partial I_C} = -2\pi f_T^2 \frac{\partial}{\partial I_C}\left(\frac{C_{eb} + C_{bc}}{g_m}\right)$$

From this relationship, the ESD current can be used to evaluate the change in the transconductance term.

What does the future bring for SiGe HBT devices and ESD? As this text is written, new discoveries in SiGe technology and ESD phenomena are occurring. Faster transistors and new structures are being developed leading to new ESD issues.

In Chapter 12, we address new directions of devices, which may or may not include the introduction of SiGe films. The next chapter will discuss FinFETs, and strained materials, and the influence that is anticipated in the ESD arena.

PROBLEMS

11.1. Derive the breakdown trigger condition using an open base common emitter configured silicon–germanium transistor necessary to protect a MOSFET gate oxide receiver network. What is the unit current gain cutoff frequency condition for a gate oxide breakdown voltage of V_{ox}?

11.2. Derive a trigger condition for a common emitter configured silicon–germanium transistor with a grounded base resistor.

11.3. Derive a trigger condition for a high-frequency unit current gain common emitter configured SiGe HBT transistor with a grounded base resistor necessary to protect a second SiGe HBT transistor with a high breakdown BVCEO and low-frequency unit current gain cutoff frequency.

11.4. Derive the dimensionless group comparing the ratio of the Wunsch–Bell power to failure of a transistor with the Johnson limit maximum power in the adiabatic regime.

11.5. Derive the dimensionless group comparing the ratio of the Wunsch–Bell power to failure of a transistor with the Johnson limit maximum power in the thermal diffusion regime.

11.6. Derive the dimensionless group comparing the ratio of the Wunsch–Bell power to failure of a transistor to the Johnson limit maximum power in the steady state regime.

11.7. Derive the dimensionless group comparing the ratio of the Wunsch–Bell power to failure of a transistor to the Johnson limit maximum power where the applied pulse width frequency is equal to the unit current gain cutoff frequency regime. Apply the Dwyer power to failure model. Which time constant would you use and what dimension would it represent?

11.8. Assuming the SiGeC and SiGe HBT devices have a unit current gain cutoff frequency values of $f*_{\mathrm{T}}$ and f_{T}, respectively, derive the breakdown condition ratio.

11.9. Assuming the SiGeC and SiGe HBT device have a unity current gain cutoff frequency values of $f*_{\mathrm{T}}$ and f_{T}, respectively, derive the breakdown ratio as a function of the Johnson limit relationship.

11.10. Given that a power clamp trigger device is required which is greater than the VDD of a power supply, but lower than the BVCEO of the high-frequency SiGe HBT device, derive the inequality for the trigger condition.

REFERENCES

1. R. L. Pritchard. Frequency response of grounded base and grounded emitter junction transistor. *Proceedings of the AIEEE Winter Meeting*, January 1954.
2. J. M. Early. Structure determined gain–band product of junction triode transistors. *Proceedings of the IRE*, December 1958.
3. J. M. Early. Maximum rapidly switchable power density in junction triodes. *IEEE Transactions on Electron Devices* 1959; **ED-6**: 322.
4. J. M. Early. Speed in semiconductor devices. *Proceedings of the IRE National Convention*, March 1962.
5. C. T. Kirk. A theory of transistor cutoff frequency (f_{t}) falloff at high current densities. *IEEE Transactions on Electron Devices* 1962; **ED-9**: 164–174.
6. E. O. Johnson. Physical limitations on frequency and power parameters of transistors. *RCA Review* June 1965; 163–177.
7. J. M. Goldey and R. M. Ryder. Are transistors approaching their maximum capabilities. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, 1963; 20.
8. H. Kroemer. Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region. Solid State Electronics 1985; **28**, 1101–1103.
9. H. Kroemer. Heterostructure bipolar transistors and integrated circuits. *Proceedings of the IEEE* 1982; **70**(1): 13–25.
10. U. Konig. SiGe and GaAs as competitive technologies for RF-applications. *Proceedings of the Bipolar Circuits and Technology Meeting (BCTM)*, Minneapolis, 1998; 87–92.
11. R. Singh, D. Harame and M. Oprysko. *Silicon Germanium: Technology, Modeling and Design*. Wiley, 2004.
12. D. Harame, J. Comfort, J. Cressler, E. Crabbe, J. Y.-C Sun and B. Meyerson. Si/SiGe epitaxial-base transistors—Part I: materials, physics, and circuits. *IEEE Transactions on Electron Devices*, 1995; **42**(3): 455–468.
13. D. Harame, J. Comfort, J. Cressler, E. Crabbe, J. Y.-C. Sun and B. Meyerson. Si/SiGe epitaxial-base transistors—Part II: process integration and analog applications. *IEEE Transaction on Electron Devices*, 1995; **42**(3): 469–482.
14. D. L. Harame, J. M. Stork, B. S. Meyerson, N. Nguyen and G. J. Scilla. Epitaxial-base transistors with ultra-high vacuum chemical vapor deposition (UHV-CVD) epitaxial/enhanced profile control for greater flexibility in device design. IEEE Electron Device Letters, 1989; **10**(4): 156–158.
15. G. L. Patton, S. S. Iyer, S. L. Delage, S. Tiwari and J. M. C. Stork. Silicon germanium base heterojunction bipolar transistors by molecular beam epitaxy. *IEEE Electron Device Letters* 1988; **9**(4): 165–167.

16. D. L. Harame, J. M. C. Stork, B. S. Meyerson, K. Y. Hsu, J. Cotte, K. A. Jenkins, J. D. Cressler, P. Restle, E. F. Crabbe, S. Subbanna, T. Tice, B. W. Scharf and J. A. Yasaitis. Optimization of SiGe HBT technology for high speed analog and mixed signal applications. *International Electron Device Meeting (IEDM) Technical Digest*, 1993; 874–876.
17. D. L. Harame, E. F. Crabbe, J. D. Cressler, J. H. Comfort, J. Y. C. Sun, S. R. Stiffler, E. Kobeda, J. N. Burghartz, M. M. Gilbert, J. A. Malinowsky and A. J. Dally. A high performance epitaxial SiGe-base ECL BiCMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1992; 19–22.
18. J. N. Burghartz, B. J. Ginsberg, S. R. Mader, T. C. Chen and D. L. Harame. Selective epitaxy base transistor (SEBT). *IEEE Electron Device Letters*, 1998; **9**(5): 259–261.
19. J. N. Burghartz, J. H. Comfort, G. L. Patton, B. S. Meyerson, J. Y-C. Sun, J. M. C. Stork, S. R. Mader, C. L. Stanis, G. J. Scilla and B. J. Ginsberg. Self-aligned SiGe-base heterojunction bipolar transistor by selective epitaxy emitter window (SEEW) technology. *IEEE Electron Device Letters* 1990; **11**(7): 288–290.
20. A. Gruhle. Prospects for 200 GHz on silicon with silicon germanium heterojunction bipolar transistors. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, 2001.
21. L. C. M. van der Over *et al*. Design of a 200 GHz SiGe HBT. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, 2001.
22. J. Kirchgessner et al. A 0.18 µm SiGe:C RF BiCMOS technology for wireless and gigabit optical communication applications. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, 2001.
23. A. Joseph. A 0.18 µm BiCMOS technology featuring a 120/100 GHz (f_T/f_{MAX}) HBT and ASIC compatible CMOS using copper interconnects. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, 2001.
24. E. J. Prinz, P. M. Garone, P. V. Schwartz, X. Xiao and J. C. Sturm. The effects of base dopant out-diffusion and undoped SiGe junction space layers in Si/SiGe/Si heterojunction bipolar transistors. *Transactions on Electronic Devices Letters* 1991; **12**: 42–44.
25. E. J. Prinz and J. C. Sturm. Analytical modeling of current gain-Early voltage products in $Si/Si_{(1-x)} - Ge_{(x)}$ heterojunction bipolar transistors. *International Electron Device Meeting (IEDM) Technical Digest*, 1991; 853–856.
26. E. F. Crabbe, J. D. Cressler, G. L. Patton, J. M. C. Stork, J. H. Comfort and J. Y. C. Sun. Current gain rolloff in graded-base SiGe heterojunction bipolar transistors. *IEEE Electron Device Letters* 1993; **14**(4): 193–195.
27. E. F. Crabbe, J. M. C. Stork, G. Baccarani, M. V. Fischetti and S. E. Laux. The impact of nonequilibrium transport on breakdown and transit time in bipolar transistors. *International Electron Device Meeting (IEDM) Technical Digest*, 1990; 463–466.
28. E. Zanoni, E. F. Crabbe, J. M. Stork, P. Pavan, G. Verzellesi, L. Vendrame and C. Canali. Extension of impact ionization multiplication coefficient measurements to high electric fields in advanced Si BJTs. *IEEE Electron Device Letters* 1993; **14**(2): 69–71.
29. J. W. Slotboom, G. Streutker, A. Pruijmboom and D. J. Gravesteijn. Parasitic energy barriers in SiGe HBTs. *IEEE Electron Device Letters* 1991; **12**(9): 486–488.
30. S. Tiwari. A new effect at high currents in hetero-structure bipolar transistors. *IEEE Electron Device Letters*, 1989; **10**: 2105.
31. J. H. Comfort, E. F. Crabbe, G. L. Patton, J. M. C. Stork, J. Y. C. Sun and B. S. Meyerson. Impact ionization reduction in SiGe HBTs. *Proceedings of the Device Research Conference (DRC)*, 1990; 21–24.
32. D. C. Wunsch and R. R. Bell. Determination of threshold voltage levels of semiconductor diodes and transistors due to pulsed voltages. *IEEE Transactions on Nuclear Science* 1968; **NS-15**(6): 244–259.
33. D. M. Tasca. Pulse power failure modes in semiconductors. *IEEE Transactions on Nuclear Science* 1970; **NS-17**(6): 346–372.

34. W. D. Brown. Semiconductor device degradation by high amplitude current pulses. *IEEE Transactions on Nuclear Science*, 1972; **NS-19**.
35. J. Whalen. The RF pulse susceptibility of UHF transistors. *IEEE Transactions on Electromagnetic Compatibility*, 1975; **EMC-17**:220–225.
36. J. Whalen and H. Domingos. Square pulse and RF pulse overstressing of UHF transistors. *Proceedings of the EOS/ESD Symposium*, 1979; 140–146.
37. D. Alexander, E. W. Enlow and R. J. Karakiewicz. Statistical variations in failure thresholds of silicon *npn* transistors subjected to electrical overstress. *IEEE Transactions on Nuclear Science*, 1980; **NS-27** (6).
38. P. L. Hower. High field phenomena and failure mechanisms in bipolar transistors. *Proceedings of the EOS/ESD Symposium*, 1980; 112–116.
39. H. Domingos. Basic considerations in electrothermal overstress in electronic components. *Proceedings of the EOS/ESD Symposium*, 1980; 206–212.
40. D. R. Alexander and E. W. Enlow. Predicting lower bounds on failure power distributions of silicon *npn* transistors. *IEEE Transactions on Nuclear Science* 1981; **NS-28** (6).
41. E. N. Enlow. Determining an emitter–base failure threshold density of *npn* transistors. *Proceedings of the EOS/ESD Symposium*, 1981; 145–150.
42. D. Pierce and R. Mason. A probabilistic estimator for bounding transistor emitter–base junction transient-induced failures. *Proceedings of the EOS/ESD Symposium*, 1982; 82–90.
43. A. L. Ward. Calculations of second breakdown in silicon diodes at microwave frequencies. *Proceedings of the EOS/ESD Symposium*, 1983; 102–107.
44. M. M. S. Hassan and H. Domingos. The double graded transistor and its beneficial effect on resistance to current mode second breakdown. *Proceedings of the EOS/ESD Symposium*, 1989; 127–135.
45. S. Voldman. The state of the art of electrostatic discharge protection: physics, technology, circuits, designs, simulation and scaling. *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 27–29 September 1998; 19–31.
46. S. Voldman, P. Juliano, R. Johnson, N. Schmidt, A. Joseph, S. Furkay, E. Rosenbaum, J. Dunn, D. L. Harame and B. Meyerson. Electrostatic discharge and high current pulse characterization of epitaxial base silicon germanium heterojunction bipolar transistors. *Proceedings of the International Reliability Physics Symposium (IRPS)*, March 2000, 310–316.
47. S. Voldman, N. Schmidt, R. Johnson., L. Lanzerotti, A. Joseph, C. Brennan, J. Dunn, D. Harame, P. Juliano, E. Rosenbaum and B. Meyerson. Electrostatic discharge characterization of epitaxial base silicon germanium heterojunction bipolar transistors. *Proceedings of the EOS/ESD Symposium*, 2000; 239-251.
48. S. Voldman et al. ESD robustness of a silicon germanium BiCMOS technology. *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, September 2000.
49. S. Voldman, L. D. Lanzerotti and R. Johnson. Emitter base junction ESD reliability of an epitaxial base silicon germanium heterojunction transistor. *Proceedings of International Physical and Failure Analysis (IPFA) of Integrated Circuits*, July 2001; 79–84.
50. S. Voldman, L. D. Lanzerotti and R. A. Johnson. Influence of process and device design on ESD sensitivity of a silicon germanium heterojunction bipolar transistor. *Proceedings of the EOS/ESD Symposium*, 2001; 364–372.
51. S. Voldman, A. Botula, D. Hui and P. Juliano. Silicon germanium heterojunction bipolar transistor ESD power clamps and the johnson limit. *Proceedings of the EOS/ESD Symposium*, 2001; 326–336.
52. S. Voldman. Variable trigger-voltage ESD power clamps for mixed voltage applications using a 120 GHz/100 GHz (f_T/f_{MAX}) silicon germanium heterojunction bipolar transistor with carbon Incorporation. *Proceedings of the EOS/ESD Symposium*, 2002; 52–61.
53. S. Voldman, B. Ronan, S. Ames, A. Van Laecke, J. Rascoe, L. Lanzerotti, and D. Sheridan. Test methods, test techniques and failure criteria for evaluation of ESD degradation of analog

and radio frequency (RF) technology. *Proceedings of the EOS/ESD Symposium*, 2002; 92–100.
54. S. Voldman, S. Strang and D. Jordan. A design system for auto-generation of ESD circuits, *International Cadence Users Group (ICUG)*, September 2002.
55. S. Voldman, S. Strang and D. Jordan. An automated electrostatic discharge computer-aided design system with the incorporation of hierarchical parameterized cells in BiCMOS analog and RF technology for mixed signal applications. *Proceedings of the EOS/ESD Symposium*, 2002; 296–305.
56. L.Lanzerotti. Suppression of boron out-diffusion in silicon germanium HBT by carbon incorporation. *International Electron Device Meeting (IEDM) Technical Digest*, 1996; 249–252.
57. N. E. B. Cowern, P. C. Zalm, P. V. D. Sluis, D. J. Graresteijn and W. B. D. Groer. Diffusion in strained SiGe. *Physical Review Letters* 1994; **72**: 2585–2588.
58. J. M. Poate, D. J. Eaglesham, G. H. Gilmer, H. J. Gossmann, M. Jaraiz, C. S. Rafferty and P. A. Stolk. Ion implantation and transient enhanced diffusion. *International Electron Device Meeting (IEDM) Technical Digest*, 1995; 77–80.
59. R. F. Lever, J. M. Bonar and A. F. W. Willoughby. Boron diffusion across silicon–silicon germanium boundaries. *Journal of Applied Physics* **83**: 1999; 1998–1994.
60. K. Rajendran, W. Schoenmaker, S. Decoutere, M. Caymax, R. Loo and W. Vandervorst. Measurement and simulation of boron diffusion in strained $Si_{1-x}Ge_x$ epitaxial layers. *IEEE Transactions on Electron Devices*.
61. K. Rajendran, D. Villaneuva, P. Moens and W. Schoenmaker. Modeling of clustering reaction and diffusion of boron in strained $Si_{1-x}Ge_x$ epitaxial layers. *Applied Physics Letters*, to be published.
62. K. Rajendran and W. Schoenmaker. Studies of boron diffusivity in strained $Si_{1-x}Ge_x$ epitaxial layers. *Journal of Applied Physics* 2001; **89**(2): 980–987.
63. K. Eberl, S. S. Iyer, S. Zalner, J. E. T sand and F. K. Lehover. Growth and strain compensation effects in the ternary $Si_{1-x-y}Ge_xC_y$ alloy system. *Applied Physics Letters* 1992; **60**: 3033–3305.
64. S. Nishikawa, A. Takeda and T. Yamaji. Reduction of transient boron diffusion in preamorphized Si by carbon incorporation. *Applied Physics Letters* 1992; **60**: 2270–2272.
65. P. A. Stolk, D. J. Eaglesham, H.-J. Gossmann and J. M. Poate. Carbon incorporation in silicon for suppressing interstitial-enhanced boron diffusion. *Applied Physics Letters* 1995; **66**, (11): 1370–1372.
66. L. Lanzerotti, J. Sturm, E. Stach, R. Hull, T. Buyuklimanli and C. Magee. Suppression of boron out-diffusion in silicon germanium HBT by carbon incorporation. *International Electron Device Meeting (IEDM) Technical Digest*, 1996; 249–252.
67. H. J. Osten, G. Lippert, G. Knoll, R. Barth, B. Heinemann, H. Rucker and P. Schley. The effect of carbon incorporation on SiGe heterojunction bipolar transistor performance and process margin. *International Electron Device Meeting (IEDM) Technical Digest*, 1997; 803–806.
68. R. Scholtz, U. Gosele, J. Y. Huh and T. Y. Tan. Carbon-induced under saturation of silicon self-interstitials. *Applied Physics Letters* 1998; **72**: 200–202.
69. H. Rucker, B. Heinemann, W. Ropke, R. Kurps, D. Kruger, G. Lippert and H. J. Osten. Suppressed diffusion of boron and carbon in carbon-rich silicon. *Applied Physics Letters* 1998; **73**: 1682–1684.
70. R. Scholz, P. Werner, U. Gosele and T. Y. Tan. The contribution of vacancies to carbon outdiffusion in silicon. *Applied Physics Letters*, 1999; **74**: 392–394.
71. H. Rucker, B. Heinemann, D. Bolze, R. Kurps, D. Kruger, G. Lippert and H. J. Osten. The impact of supersaturated carbon on transient enhanced diffusion. *Applied Physics Letters*, 1999; **74**: 3377–3379.
72. H. Kurata, K. Suzuki, T. Futasugi and N. Yokoyama. Shallow p-type SiGeC layers synthesized by ion implantation of Ge, C, and B in Si. *Applied Physics Letters* 1999; **75**: 1568–1570.

73. H. Rucker and B. Heinemann. Tailoring dopant diffusion for advanced SiGe:C heterojunction bipolar transistor. *Solid State Electronics* 2000; **44**: 783–789.
74. K. Rajendran and W. Schoenmaker. Modeling of complete suppression of boron out-diffusion in $Si_{1-x}Ge_x$ by carbon incorporation. *Solid State Electronics*, 2001; **45**: 229–233.
75. B. Ronan, S. Voldman, L. Lanzerotti, J. Rascoe, D. Sheridan and K. Rajendran. High current transmission line pulse (TLP) and ESD characterization of a silicon germanium heterojunction bipolar transistor with carbon incorporation. *Proceedings of the International Reliability Physics Symposium (IRPS)*, Dallas, April 2002; 175–183.

12 Nanostructures and ESD

In the bipolar technology, the dominant device has changed from the silicon bipolar junction transistor to the silicon–germanium heterojunction bipolar transistor and is presently, changing to the silicon–germanium–carbon HBT device. In CMOS technology, planar MOSFETs dominate the bulk and silicon-on-insulator (SOI) technology. Continued interest exists in extending the life of the MOSFET transistor by addressing dual-gate SOI, and ultra-thin film SOI (UT-SOI) technologies. In the magnetic recording industry, the thin-film giant magnetoresistive (GMR) device is the dominant device with future interest in the tunneling magnetoresistive (TMR) device.

But the continued interest in the devices of the future has captured the imagination of physicists, chemists, and engineers. As the structures are scaled to nanometer dimensions, the interest in nanostructures, and quantum devices will continue. Nanostructures have increased interest in the semiconductor industry as engineers anticipate obstacles as the miniaturization of semiconductor devices.

In this chapter, we will first discuss the directions of MOSFETs from two-dimensional MOSFET devices to the three-dimensional MOSFET trend [1–10]. We will discuss the continued interest in narrow silicon pillars, where today's focus is on the FinFET device [6,8,9]. Second, a new direction in MOSFET structures is the strained silicon MOSFETs. Strained silicon technology uses new materials to modify the silicon lattice [11–27]. Third, we will discuss nanotubes [28–44]. A radical departure is taking place from the traditional processing of semiconductor devices. Nanowire, nanotube, and nanobelt research and development continues to make progress in understanding the material, properties and devices of the future. What will be the electrostatic discharge sensitivity of these devices? How will we protect them from static charge? What ESD devices can we construct from these new devices in the future? Let's take a look into the future.

12.1 FinFETs

With the continued struggle to scale MOSFET devices to the sub-tenth micrometer era, semiconductor engineers have been pursuing new directions in MOSFETs [1–6]. To

ESD: Physics and Devices Steven H. Voldman
 ISBN: 0-470-84753-0

achieve both high density and high performance, the MOSFET is leaving the paradigm of evolution in two dimensions and must address a new revolutionary move to the third dimension. A potential evolutionary path for the MOSFET was a two-dimensional bulk CMOS planar MOSFET device, to the two-dimensional single-gate (SG) silicon-on-insulator (SOI) MOSFET, to the two-dimensional dual-gate (DG) SOI MOSFET device. The dual-gate SOI MOSFET aims to provide a means to produce more current for a given planar device. The problem with the DG SOI MOSFET is processing costs and alignment of the second gate. For this reason, new directions have been taken toward 'surround' gate or 'wrap-around' gate structures [1–5]. In 1986, Takahashi *et al.* [1] proposed the surround gate transistor (SGT) device with the objective of achieving a smaller transistor structure. D. Hisamoto *et al.* [2,3,5] proposed the fully depleted lean channel transistor (DELTA) device which was a novel vertical ultra-thin SOI MOSFET structure. This evolution has progressed toward a silicon pillar device with a wrap-around gate in both bulk CMOS and in SOI technology. Tang *et al.* [8] developed a quasi-planar double-gate device known as a 'FinFET.' Concepts of surround gates, wrap-around gates and nonplanar dual gates were all different strategies on constructing the nonplanar MOSFET in narrow width silicon pillars, leaving the wafer surface to form the three-dimensional MOSFET structures. Figure 12.1 is an example of a *n*-channel MOSFET FinFET structure.

In a FinFET structure [9], the key design parameters are the fin height H, fin thickness T_{si}, the effective channel length L_{eff}, and the number of parallel fin structures N_{Fin}. For analysis of the FinFET, we can define an effective channel length as

$$L_{eff} = L_{poly} - \Delta L$$

In a double-gated (DG) FinFET device, the contours merge, at

$$T_{Si} \cong -2\frac{\varepsilon_{Si}}{\varepsilon_{ox}} T_{ox}$$

and we can define an effective film thickness as

$$T_{eff} = T_{Si} + 2\frac{\varepsilon_{Si}}{\varepsilon_{ox}} T_{ox}$$

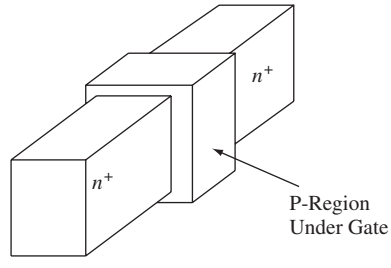

Figure 12.1 *n*-Channel Fin-FET

It has been shown by Suzuki that a more accurate solution is [10]

$$T_{\text{eff}} = \sqrt{T_{\text{Si}}^2 + 4\frac{\varepsilon_{\text{Si}}}{\varepsilon_{\text{ox}}} T_{\text{ox}} T_{\text{Si}}}$$

The FinFET structure forms a parallelepiped surrounded by isolation material below and above the conducting region, and the thermal sinks formed by the boundary conditions. For evaluation of a parallelepiped in an infinite insulating medium let us assume a source has the dimension of a FinFET width $W = T_{\text{eff}}$ in the x-direction, and L_{eff} in the y-direction, and FinFET height H in the z-direction, but applying a parallelepiped displaced by a distance D below the boundary condition $z = 0$ and an image source of equal and opposite strength above the $z = 0$ plane at $z = \text{D}$.

$$T(x, y, z; t) = \frac{1}{8\rho c V(\kappa) 3/2} \int_{t'=0}^{t'=t} \frac{P(t')\mathrm{d}t'}{[(t - t')]^{3/2}} F(x - x', y - y', z - z', t - t')$$

with

$$F(x - x', y - y', z - z', t - t') = F_x(x - x' : t - t') F_y(y - y'; t - t') F_z(z - z'; t - t')$$

where

$$F_x(x - x', t - t') = \int_{-W/2}^{W/2} \exp\left\{ -\frac{(x - x')2}{4\kappa(t - t')} \right\} \frac{\mathrm{d}x'}{\sqrt{\pi}}$$

$$F_y(y - y', t - t') = \int_{-L/2}^{L/2} \exp\left\{ -\frac{(y - y')2}{4\kappa(t - t')} \right\} \frac{\mathrm{d}y'}{\sqrt{\pi}}$$

$$F_z(z - z', t - t') = \int_{-(D+H)}^{-D} \exp\left\{ -\frac{(z - z')2}{4\kappa(t - t')} \right\} \frac{\mathrm{d}z'}{\sqrt{\pi}} + \int_{D}^{D+H} \exp\left\{ -\frac{(z - z')2}{4\kappa(t - t')} \right\} \frac{\mathrm{d}z'}{\sqrt{\pi}}$$

The integral can be expressed as an error functions by a transformation of variables. The expression for temperature in an infinite medium can be put in the form

$$T(x, y, z; t) = \frac{1}{8C} \int_{t'=0}^{t'=t} P(t') H(x, y, z; t - t') \mathrm{d}t'$$

where $V = L_{\text{eff}} T_{\text{eff}} H$ is the volume of a single FinFET region, and letting $C = \text{c}\rho V$, we can write the function containing the spatial dependence as the product of the error functions

$$H(x, y, z; t - t') = H(z; t - t') \prod_{i=x,y} \left[erf\left(\frac{(L_{x_i}/2) + x_i}{\sqrt{4\kappa(t - t')}} \right) + erf\left(\frac{(L_{x_i}/2) - x_i}{\sqrt{4\kappa(t - t')}} \right) \right]$$

and

$$H(z;t-t') = \left[erf\left(\frac{z+D+H}{\sqrt{4\kappa(t-t')}}\right) + erf\left(\frac{-D-z}{\sqrt{4\kappa(t-t')}}\right) + erf\left(\frac{z-D}{\sqrt{4\kappa(t-t')}}\right) + erf\left(\frac{D+H-z}{\sqrt{4\kappa(t-t')}}\right)\right]$$

From this solution, assuming the FinFET is surrounded by insulating regions, the solution for self-heating within the FinFET can be obtained.

In the case of a plurality of FinFETs in parallel, the method of images can be used in the dimension of the FinFET widths (x-direction) where each are a thermal source. The complete solution can be obtained by assuming a row of spaced thermal sources at a given temperature.

For ESD phenomena, there are two issues. The first issue is how the current distributes between multiple parallel FinFET devices. A second issue is the relative width of a current constriction to the effective fin width and height. The first issue is similar to any other parallel configuration in that the current distribution is a function of the ballasting and matching between any two parallel elements during second breakdown. The second issue is the more interesting. Will the width of the FinFET be smaller than the current constriction width, forcing the current to flow through multiple elements? Will the full width of the FinFET be utilized during an ESD event having the 'effective width' during an ESD event equal to the FinFET equivalent width ($W_{\text{eff}} = 2H$)? Will the constriction fill the full width of the device T_{eff}? Hence the volumetric nature of the current constriction and relative scale length compared with the dimension fin height H, and fin width T_{eff} will be of interest.

From our earlier analysis of current constriction in planar MOSFETs we can anticipate that the relationship may have a one-to-one equivalency of planar MOSFET width to the fin height, of N parallel FinFETs, assuming that there is conduction in some number M where M is equal to or less than N

$$W_{\text{eff}} \simeq \sum_{i=1}^{i=N} (2H_{\text{eff}})_i = \sum_{i=1}^{i=N} \frac{(\mathrm{I_m})_i}{\displaystyle\int_{T_{\max}}^{T} \mathrm{d}T \left[\frac{4\rho K}{\sqrt{2\left\{\int_{T_{\max}}^{T} \rho(T')K(T')\mathrm{d}T'\right\}}}\right]}$$

When the fin height and width are of the same order of magnitude, it is possible that the arithmetic mean may serve as a better metric, where the H_{eff}, and the T_{eff} are the ESD effective widths based on the percentage of the fin height and the electrical effective FinFET width.

$$W_{\text{eff}} \simeq \sqrt{H_{\text{eff}} T_{\text{eff}}} = \sum_{i=1}^{i=N} \frac{(\mathrm{I_m})_i}{\displaystyle\int_{T_{\max}}^{T} \mathrm{d}T \left[\frac{4\rho K}{\sqrt{2\left\{\int_{T_{\max}}^{T} \rho(T')K(T')\mathrm{d}T'\right\}}}\right]}$$

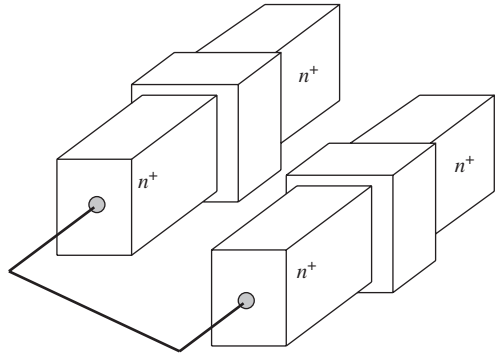

Figure 12.2 *n*-Channel FinFET parallel elements

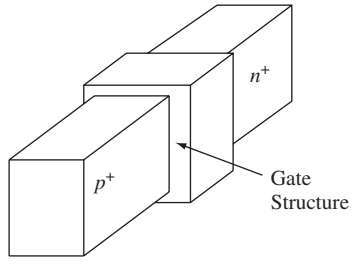

Figure 12.3 p^+/n^+ Fin diode

For functional devices or ESD applications, the *n*-channel device is made into a plurality of parallel elements. Figure 12.2 is an example of parallel FinFET structures.

For ESD applications, the two-dimensional polysilicon-bound gated diode structure can be extended into a CMOS FinFET technology. This can be constructed by forming the anode, using the *p*-channel FinFET process, and the cathode, using the *n*-channel FinFET process. The mask is placed over the FinFET gate to form this p^+/n^+ Fin diode (Figure 12.3).

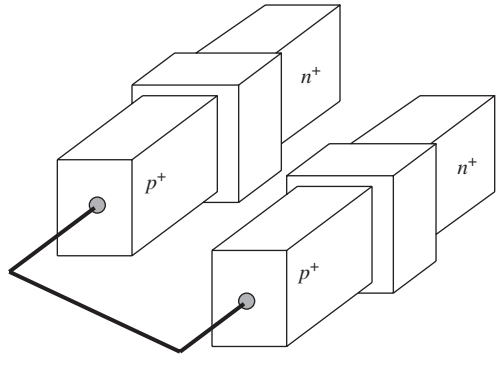

Figure 12.4 Fin diode parallel elements

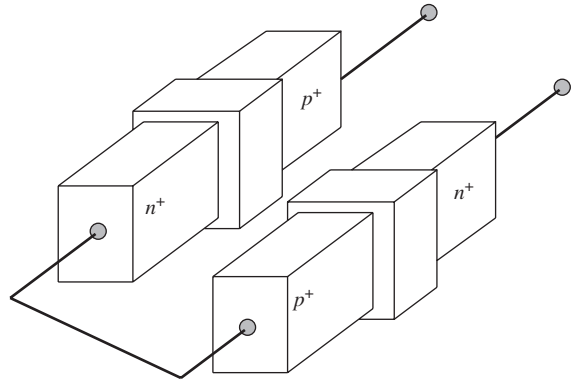

Figure 12.5 Fin diode series elements

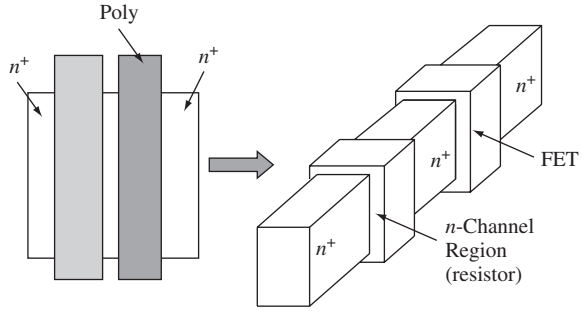

Figure 12.6 Resistor-ballasted FinFeT structure

To provide ESD protection, the ESD structure will be required to be formed in a plurality of elements in parallel. Figures 12.4 and 12.5 show the elements in parallel and series configurations.

Mapping bulk CMOS techniques into a FinFET technology can be realized by forming resistor ballasted FinFET elements. Figure 12.6 shows an example on how a FinFET can add a resistor element, using a gate structure in series with the source or drain structure.

12.2 STRAINED SILICON DEVICES AND ESD

Strained devices are used for future applications to provide improvement in the speed of bipolar and MOSFET transistors [11–23]. Strain changes the band structure of a material, leading to changes in the bandgap, the impact ionization, resistance and mobility. These will influence both the low-current and high-current characteristics of a device. As a result, the response of a strained device to an electrostatic discharge event will differ from the standard unstrained material response. The fractional change of resistance due to mechanical stress can be expressed as the sum over all the indices of the piezoresistance tensor and the stress tensor [11]

$$\frac{\Delta\rho\omega}{\rho} = \sum_{\lambda=1}^{\lambda=6} \pi_{\omega\lambda} X_\lambda$$

For cubic symmetry, the piezoresistance tensor is expressed as

$$\pi_{\omega\lambda} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{11} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{12} & \pi_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi_{44} \end{bmatrix}$$

Conductivity can be expressed as a tensor where the conductivity is the sum of the carriers near the band extremum

$$\sigma_{\alpha\beta} = \sum\nolimits_i \sigma^i_{\alpha\beta}$$

and

$$\frac{\Delta\sigma_\omega}{\sigma} = \sum_{\lambda=1}^{\lambda=6} \pi_{\omega\lambda} X_\lambda$$

From solid state physics, we can relate the conductivity of an extremum to the integral of the derivative of the Fermi distribution, the relaxation time tensor, and the group velocity of the carriers

$$\sigma^i_{\alpha\beta} = -q^2 \int \frac{\partial f}{\partial E} \tau^i_{\alpha\beta} \nu^i_\alpha \nu^i_\beta \, \mathrm{d}K$$

As the strain of a material increases, the conductivity is modulated by the changes in the group velocity of carriers, and the relaxation time tensor. The relaxation time tensor can be assumed to be isotropic and a function of energy only. As strain modifies the conductivity and the resistivity, the current and voltage characteristics of the device change.

$$f_0 = \frac{1}{1 + \exp\left(\dfrac{E_i - E_\mathrm{F}}{kT}\right)}$$

$$\Delta\nu^i_\alpha = \frac{1}{\hbar} \frac{\partial E_i}{\partial K_\alpha}$$

The relaxation time can be simplified by having it a function of energy only, and simplified to the form

$$\Delta\tau_i = \frac{\partial \tau_i}{\partial E}\Delta E_i$$

$$\tau = \tau_0 E^s$$

In this form, it can be shown that the conductivity can be expressed as the ratio of the derivative of the Fermi integral over the Fermi integral evaluated as a function of E_F/kT ($F(E_F/kT)$)

$$\frac{\Delta\sigma}{\sigma} \simeq \frac{1}{kT}\frac{1}{\left(s+\frac{3}{2}\right)}\left[\frac{F'_{(s+1/2)}}{F_{(s+1/2)}}\right]$$

The piezoresistance coefficient $P(N,T)$ is a function of both doping concentration and temperature. This can be expressed as

$$P(N,T) = \frac{300}{T}\left[\frac{F'_{(s+1/2)}}{F_{(s+1/2)}}\right]$$

where

$$\Pi(N,T) = P(N,T)\Pi(300\,\mathrm{K})$$

Of interest to the ESD protection, as the strained material temperature increases, the piezoresistance coefficient P(N,T) varies according to the above relationship. The change of conductivity is related to the strain in the material. This can be obtained from evaluation of the conductivity of each ellipsoid, and evaluation of the change in the number of electron carriers in that ellipsoid. The assumption is the mobility in the ellipsoid is not a function of the stress [11]

$$\Delta\sigma^i = q\mu^i\Delta n^i$$

where

$$\frac{\Delta n^i}{n^i} = \frac{1}{kT}\left(\Delta E^i - \Delta E_F\right)$$

$$\Delta E_F = \frac{1}{\nu}\sum_i \Delta E^i$$

The change in the energy levels of the individual ellipsoid can be related to the strain tensor, and the associated shift in the band structure. For ESD events, the strain can be the initial mechanical strain from the lattice strain as well as the thermal strain from self-heating. From stress–strain theory, if the material is Hookean, in a linear region linear

superposition applies and the total strain is the sum of the initial strain at room temperature and the thermal strain. During an ESD event, the initial strain and the thermal strain must be addressed to anticipate the total response of the strained semiconductor device.

With the associated strain in the lattice structure, other physical parameters also are modified. As the bandgap is modified, the electron impact ionization rate is also modified. As discussed in Chapter 1, as carriers are accelerated in a medium, energy is transferred from the electric field to the carriers. As the electric field increases, carriers approach a limiting drift velocity, and further increases lead to thermal vibrations. As carriers are accelerated, there is a competition between energy transmitted to the electron and energy transmitted to the lattice. The probability of ionizing a carrier can be expressed as

$$e^{-\{\varepsilon_i/qEl_r\}}$$

where the probability is a function of the ionization threshold, the energy of the carrier, and mean free path of optical phonon scattering. Above the threshold for ionization, the energy balance is

$$qE = \alpha_i \varepsilon_i + \alpha_r \varepsilon_r$$

where the LHS is the energy of the carrier, and the RHS is the energy transfer to impact ionization and the energy transfer to optical phonon generation. The impact ionization coefficient is the number of ionizing events divided by distance. From this we can establish a relationship of the probability that impact ionization can occur compared with phonon generation, when the energy of the carrier is above the impact ionization energy. The electron impact ionization rate can be expressed as a function of an ionization rate factor α_{n0}, the mean free path, and the electric field in the direction of the current flow (e.g. ith direction).

$$\alpha_n = \alpha_{n0} \exp\left\{-\frac{E_g(T)}{q\lambda_n E_n^i}\right\}$$

The mean free path in the direction of interest is related to the energy relaxation time in that same direction.

Strain can be introduced by introduction of a mechanical force, or by placement of atoms which do not naturally fit into the silicon lattice. For example, as discussed in the chapter on silicon–germanium, germanium modulates the bandgap and band structure. With a position-dependent Ge concentration (e.g. x-direction), the intrinsic carrier concentration can be expressed as

$$n_i^2(x) = \gamma n_{i0}^2 \, \mathrm{e}^{\Delta E(x)/kT}$$

and

$$\Delta E_{\mathrm{g}}(x) = \Delta E_{\mathrm{g,b}}^{\mathrm{app}} + \Delta E_{\mathrm{g,Ge}}(\mathrm{grade})(x/W_{\mathrm{b}}kT) + \Delta E_{\mathrm{g,Ge}}^{(0)}$$

where $\Delta E_g(x)$ is the position-dependent bandgap; $\Delta E_{g,b}^{app}$ is the heavy doping-induced apparent bandgap narrowing; and $\Delta E_{g,Ge}$ (grade) is the difference between the band bending at the base–collector junction; $\Delta E_g(x = W_b)$, and the band bending at the base–emitter junction, ΔE_g $(x = 0)$. The effective density of states ratio of SiGe and Si is expressed as

$$\gamma = (N_C N_V)_{SiGe}/(N_C N_V)_{Si}$$

The bandgap energy of strained silicon ($E_{g,SS}$) can be expressed as [26,27]

$$E_{g,SS} = 1.08 - 0.4x$$

where x is the Ge mole fraction in the Si_{1-x} Ge_x region. The changes in the drain current, temperature-dependent carrier mobility, and bandgap all play a role in the response as the structure undergoes self-heating during an ESD event. From our results, the energy relaxation time will increase with increased strain (e.g. increased mole fraction of germanium) at higher temperatures. Choi *et al.* noted that as a result of the change in the impact ionization rate (e.g. changes in the directional components), the avalanche breakdown will differ in a strained Si MOSFET. During an ESD event, the parasitic bipolar transistor will turn on at a lower trigger voltage, allowing for a low-voltage first and second trigger condition [27]. Hence, the introduction of strain, either mechanically or introduction of Ge into a Si lattice leads to modulation of the high-current characteristic of a strained Si MOSFET device compared to unstrained Si MOSFET devices. This will have an influence on the ESD robustness as strained silicon devices are used in semiconductor products.

12.3 NANOTUBES AND ESD

Nanotubes and nanostructures are of growing interest with materials scientists, and engineerings [28–44]. The use of nanotubes as electronic components opens the question of the ESD nature of these elements. What will be the ESD sensitivity of nanotubes? Will we need to protect them from static charge? What ESD devices can we construct from nanotube structure, in the future? A nanowire (NW) of considerable interest is the carbon nanotube (CNT). Carbon nanotubes are cylindrical strands of carbon which conduct current along the axis of the cylinder. Carbon nanotubes can be classified as single-wall nanotubes (SWNT) and multi-wall nanotubes (MWNT) devices. The electrical nature of the nanotubes can be metallic or semiconductor [28–32]. Figure 12.7 is an example of a carbon nanotube.

Resistors, diodes, bipolar transistors and field effect transistors can be constructed from nanowires by physically crossing nanowires, which allows conduction between the two NW elements [29]. By placing two nanotubes in a crossed configuration, a resistor can be formed between two *n*-type or two *p*-type nanotubes. This forms an *nn* resistor, and a *pp* resistor element. A cross-wired *nn* or *pp* resistor element (Figure 12.8) have a linear *I–V* characteristic at low currents.

By crossing a *p*- and an *n*-type NW, a *pn* NW diode structure can be formed (Figure 12.9). In a forward-bias mode of operation, current flows from the *p*-NW to

Figure 12.7 Carbon nanotubes (CNT)

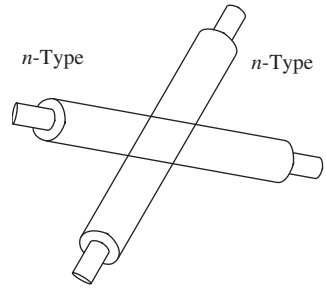

Figure 12.8 Crossed nanowire *nn* junction

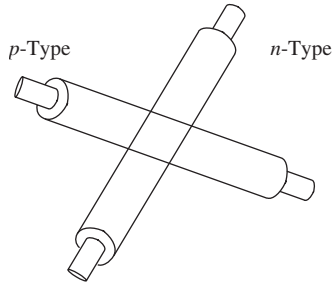

Figure 12.9 Crossed nanowire *pn* junction

the *n*-NW at the location where the two nanowires intersect. Nanowire bipolar transistors have been constructed by the coupling of two *pn* NW diode elements. A NW *npn* transistor can be constructed by using a *p*-type NW which crosses two *n*-type NW elements. In the same fashion as the nanowire diode structure, two intersections are present, which allows the current to flow from one nanowire to the second.

The nanowire field effect transistor (NW-FET) is formed by the crossing of two nanowires (Figure 12.11). The crossed NW FET (cNW-FET) is also formed by crossing two nanowires. At least one of the nanowires has an oxide coating. Using one nanowire as the source-to-drain wire, current is conducted along the nanowire to establish the source-to-drain current flow. The second nanowire crosses the first nanowire. By biasing the second nanowire, a gating action occurs, preventing conduction from the source-to-drain regions along the first nanowire. From these cross-over structures, future electronic devices will be constructed [29].

To understand the current and voltage characteristics of the nanostructures, we can evaluate a system assuming a two-dimensional quantum well in an x- and y-direction with a zero potential for $0 < x < L_x$, and $0 < y < L_y$, and infinite outside of this region.

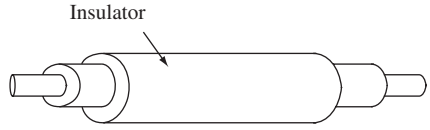

Figure 12.10 Carbon nanowire with insulator

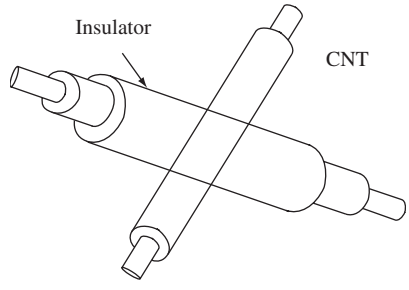

Figure 12.11 Carbon nanowire field effect transistor

From quantum mechanics, the confinement of the electron to this region leads to an energy spectrum of form

$$E = E_{x,y} + \frac{\hbar^2 k_z^2}{2m}$$

and

$$E_{x,y} = \frac{\hbar^2}{2m}\left[\left(\frac{n_x\pi}{L_x}\right)^2 + \left(\frac{n_y\pi}{L_y}\right)^2\right]$$

where n_x and n_y are integers. For quantum wires, we can assume that the device length is much greater than the width. Letting $L_y << L_x$, we can assume for the y-direction that the electron is in the lowest energy state, and the only state is $n_y = 1$. The energy spectrum reduces to the case of

$$E = \frac{\hbar^2}{2m}\left[\left(\frac{n_x\pi}{L_x}\right)^2 + \left(\frac{\pi}{L_y}\right)^2\right] + \frac{\hbar^2 k_z^2}{2m}$$

The electrical current for the two-dimensional system is [29]

$$I = q\sum\nolimits_{n_x} v_{nx}(E_F) D_{nx}(E_F)\Delta\mu$$

where $v_{nx}(E_F)$ is the Fermi velocity of electrons in the band n_x, $D_{nx}(E_F)$ is the density of states function at the Fermi energy, and q is the electron charge, and $\Delta\mu$ is the electrochemical potential between the two electrical connections on the left- and right-hand

side. The summation of the sub-bands is restricted to the energy levels between the Fermi level and the sum of the Fermi level and the electro-chemical potential.

For one-dimensional systems, the product of the Fermi velocity and density of states function can be reduced to

$$v(E) = \sqrt{\frac{2\sqrt{E}}{m}}$$

and

$$D(E) = \sqrt{\frac{2m}{h\sqrt{E}}}$$

The current equation can then be simplified and expressed as step functions associated with the energy levels [43]

$$I = \frac{2q}{h}\Delta\mu \sum_{n_x} H(E - E_{n_x})H(E_{n_x} - E_{\mathrm{F}})$$

where

$$E = E_F + \Delta\mu$$

which can be expressed as

$$I = \frac{2q}{h}\Delta\mu N_{\mathrm{C}}$$

$$N_{\mathrm{C}} = \sum_{n_x} H(E - E_{n_x})H(E_{n_x} - E_{\mathrm{F}})$$

where N_{C} is the total number of propagating channels. The conductance can be solved for by dividing by the external bias potential. The external bias potential is related to the electrochemical potential as $\Delta\mu = qV$. We can express the conductance as [44]

$$G = \frac{I}{V} = \frac{2q^2}{h}N_{\mathrm{C}} = G_0 N_{\mathrm{C}}$$

where the quantum conductance can be expressed as

$$G_0 = \frac{2q^2}{h}$$

This expression assumes no scattering. Hence, it is apparent that this formalism can be extended assuming a transmission probability T which we will express as

$$G = \frac{2q^2}{h}T$$

Buttiker [44] extended the development to include the contact resistance effects. As in transmission line theory, the boundary conditions influence the transmission and reflection characteristics, in quantum devices, the quantum mismatch of the wavefunctions at the boundaries of the quantum wire nano-structure influence the resistance of the structure.

The Buttiker–Landauer formulation [43,44] includes the multi-probe case addressing the two transmission characteristics at the boundaries, and the nonlinear current–voltage relationship. The current–voltage characteristic can be expressed as

$$I = \frac{2q^2}{h} \int_{\Delta_1}^{\Delta_2} \mathrm{d}ET(E, V)[f_l(E) - f_r(E)]$$

where the transmission coefficient is a function of the energy and bias voltage, and the last term is the difference between the Fermi distribution function at the left- and right-hand side of the device at the contacts (e.g. the reservoirs of electrons).

The conductance of the nanowire structures is a function of the number of parallel conductance paths that exist in the nanostructure. Collins *et al.* [30] showed that the nanowire *I–V* characteristic is a function of the number of cylindrical concentric shells. In a carbon multi-wire nanotube (MWNT) device, the total conductance is a step-wise conductance. The conductance of these MWNT devices can be engineered using electrical breakdown. As the electrical breakdown is achieved, the total current is reduced in quantized steps leading to a reduced current magnitude. From the current magnitude, the number of cylindrical concentric shells in the carbon MWNT bundle can be determined.

The work of Collins *et al.* [30,31] showed that electromigration does not occur in the carbon MWNT, but instead the step-wise conductance is evident. The carbon MWNT structures have current carrying capability of $10\,\mu A/nm^2$. When the current level is exceeded, the carbon MWNT structure has an electrical breakdown leading to the step-wise current conductance.

From an ESD perspective, the nature of the MWNT leads to interesting new ESD devices, and failure criteria. For example, the current carrying capability and design of ESD networks will be determined by the number of shells needed to carry a given current magnitude. ESD design will consist of using crossed structures such as crossed *nn* resistors, crossed *pn* and other elements discussed in parallel or series configurations.

From an ESD failure viewpoint, the failure will be determined by the change in the physical structure shell number or total current carrying capability. Given a circuit with an *n*-shell MWNT structure, when electrical overstress is evident, there will be a step-wise reduction in the current conductance. Dependent on the current specification bounds of a given circuit, the failure criteria will be classified from the evidence of the number of shells that still exist or the reduction of the conductance.

As these new nanostructures replace present-day elements, ESD and the elements sensitivity to ESD will be a subject of considerable interest as a result of the unique current and voltage characteristics in the single and multi-walled nanotubes. These new structures will influence ESD design, circuits, and failure criteria.

12.4 THE FUTURE

Electrostatic phenomena have been a curiosity from early times. Approximately two millennia and six hundred years have past since the first awareness of electrostatic attraction by Thales of Miletus. With the growth and miniaturization of electronics, the necessity of electrostatic phenomena and understanding in nanoelectronics and nanostructures, will increase. Although we can synthesize new materials and build new processes, will we be able to construct nanostructures and make them reliable and insensitive to electrostatic phenomenon? Where are we going in the future? Hopefully, this text has opened the door to a higher understanding in this expanding field for the future, and will be a new beginning for this millennium.

PROBLEMS

12.1. Derive the number of parallel FinFETs needed to achieve a 2 kV HBM protection level assuming a current level of 1.2 A.
12.2. Derive the number of layers a carbon nanotube must have in order to discharge 1 A of current.
12.3. Demonstrate an ESD protection strategy using *pn* cross-wiring for a carbon nanotube FET.

REFERENCES

1. H. Takahashi, K. Sunouchi, N. Okabe, A. Nitayama, S. Hieda, P. Horiguchi and F. Masuaoka. High performance CMOS surrounding gate transistor (SGT) for ultra high density LSI. *International Electron Device Meeting (IEDM) Technical Digest*, 1986; 222–225.
2. D. Hisamoto, T. Kaga, Y. Kawamoto and E. Takeda. A fully depleted lean-channel transistor (DELTA)—a novel vertical ultra-thin SOI MOSFET. *International Electron Device Meeting (IEDM) Technical Digest*, 1989; 833–836.
3. D. Hisamoto *et al.* Impact of the vertical SOI DELTA structure on planar device technology. *IEEE Transaction on Electron Devices*, 1991; **ED-38**(6): 1419–1424.
4. C. P. Auth and J. D. Plummer. Vertical fully-depleted surrounding gate MOSFETs on sub-0.1 μm thick silicon pillars. *Device Research Conference (DRC) Technical Digest*, 1996.
5. D. Hisamoto, W. C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T. King, J. Bokor and C. Hu. A folded channel MOSFET for deep sub-tenth micron era. *International Electron Device Meeting (IEDM) Technical Digest*, 1998; 1032–1035.
6. X. Huang, W. C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. K. Choi, K. Asano, V. Subramanian, T. King, J. Bokor and C. Hu. Sub 50-nm FinFET: PMOS. *International Electron Device Meeting (IEDM) Technical Digest* 1999.
7. Y. K. Choi *et al.* Ultra-thin body SOI MOSFETs for deep sub-tenth micron era. *IEEE Electron Device Letters*, **21**(5): 2000.
8. S. Tang, L. Chang, N. Lindert, Y. K. Choi, W. C. Lee, X. Huang, V. Subramanian, J. Bokor, T. King and C. Hu. FINFET—a quasi-planar double-gate MOSFET. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, 2001; 118–119.

9. J. Kedzierski *et al*. High performance symmetric-gate and CMOS compatible V_t asymmetric gate FinFET devices. *International Electron Device Meeting (IEDM) Technical Digest*, 2002; 437–440.
10. K. Suzuki *et al*. *IEEE Transactions on Electron Devices*, 1996; **ED-43**(7): 1166.
11. Y. Kanda. A graphical representation of piezoresistance coefficients in silicon. *IEEE Transaction on Electron Devices*, 1982; **ED-29**(1): 64–71.
12. J. L. Egley and D. Chidambarrao. Strain effects on device characteristics: implementation in drift-diffusion simulators. *Solid Sate Electronics* 1993; **36**(12): 1653–1664.
13. Akemi Hamada *et al*. A new Aspect of mechanical stress effects in scaled MOS devices. *IEEE Transactions on Electron Devices* 1991; **ED-38**(4): 895–900.
14. Shinya Ito *et al*. Mechanical stress effect of etch-stop nitride and its impact on deep submicron transisitor design. *International Electron Device Meeting (IEDM) Technical Digest*, 2000; 247–250.
15. F. Ootsuka *et al*. A highly dense, high-performance 130 nm node CMOS technology for large-scale system-on-a-chip applications. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 2000; 575–578.
16. A. Shimizu *et al*. Local mechanical-stress control (LMC): a new technique for CMOS-performance enhancement. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 2000.
17. A. Steegen *et al*. Silicide-induced pattern density and orientation dependent transconductance in MOS transistors. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 1999; 497–500.
18. H. Rueda *et al*. Mechanical stress characterization of shallow trench isolation by Kelvin probe force microscopy. *Materials Research Society Symposium Proceedings*, 1999; **568**, 245.
19. G. Scott *et al*. NMOS drive current reduction caused by transistor layout and trench isolation induced stress. *Proceedings of the International Electron Device Meeting (IEDM) Technical Digest*, 1999; 827–830.
20. T. Hook *et al*. The dependence of channel length on channel width in narrow-channel CMOS devices for 0.35–0.13 μm technologies. *IEEE Electron Device Letters* 2000; **21**(2): 85–87.
21. J. Thurn and R. F. Cook. Stress hysteresis during thermal cycle of plasma-enhanced chemical vapor deposited silicon oxide films. *Journal Applied Physics*, 2002; **91**(4): 1988–1992.
22. T. Ohzone *et al*. Electrical characteristics of $0°/+/-45°/90°$ orientation CMOSFET with source/drain fabricated by various ion-implantation methods. *Proceedings of the IEEE International Conference on Microelectronic Test Structures (ICMTS)*, 1998; **11**: 133–138.
23. M. Mayer *et al*. Complete set of piezo-resistive coefficients of CMOS n(+)-diffusion. *Journal of Micromechanics Microengineering*, 1998; **8**, 158–160.
24. R. C. Jaeger *et al*. CMOS stress sensor circuits using piezo-resistive field effect transistors (PIFETs). *Symposium on VLSI Circuits Digest of Technical Papers*, 8–10 June 1995; 4344.
25. A. P. Dorey and T. S. Maddern. The effect of strain on MOS transistors. *Solid State Electronics*, 1969; **12**(4): 185–189.
26. K. Rim, J. L. Hoyt and J. F. Gibboms. Fabrication and analysis of deep submicron strained Si *n*-MOSFETs. *IEEE Transactions on Electron Devices* 2000; **ED-47**: 1406–1415.
27. C. H. Choi, J. H. Chun and R. Dutton. Electro-thermal characteristics of strained Si MOSFETs in high current operation. *IEEE Transactions on Electron Devices* 2004.
28. M. Dresselhaus, G. Dresselhaus and P. Avouris. *Carbon Nanotubes: Synthesis, Structure Properties, and Applications*. Berlin: Springer, 2001.
29. Z. L. Wang. *Nanowires and Nanobelts: Materials, Properties and Devices, Vol 1: Metal and Semiconductor Nanowires*. Dordrecht: Kluwer, 2003.
30. P. G. Collins, M. S. Arnold and P. Avouris. Engineering carbon nanotubes and nanotube circuits using electrical breakdown. *Science*, 2001; **292**: 706.

31. P. G. Collins, M. Hersam, M. Arnold, R. Martel and P. Avouris. Current saturation and electrical breakdown in multi-walled carbon nanotubes. *Physical Review Letters* 2001; **86**(14): 3128.
32. A. Rochefort, M. Di Ventra and P. Avouris. Switching behavior of semiconductor carbon nanotubes under an external electric field. *Applied Physics Letters* 2001; **78**(17): 2521.
33. K. Liu, P. Avouris, R. Martel and W. K. Hsu. Electrical transport in doped multi-walled carbon nanotubes. *Physical Review B* 2001; **63**.
34. U. Landmann, R. Barnett, A. G. Scherbakov and P. Avouris. Metal–semiconductor nanocontacts: silicon nanowires. *Physical Review Letters* 2001; **85**(9): 1958.
35. A. Rochefort, P. Avouris, F. Lesage and D. Salahub. Electrical and mechanical properties of distorted carbon nanotubes. *Physical Review B, November* 1999.
36. H. Stahl. J. Appenzeller, R. Martel, P. Avouris and B. Lengeler. Intertube coupling in ropes of single wall carbon nanotubes. *Physical Review Letters* 2000; **85**(26): 5186.
37. H. R. Shea, R. Martel and P. Avouris. Electrical transport in rings of single wall nanotubes: one-dimensional localization. *Physical Review Letters*, 2000; **84**(19): 4441.
38. N. Lang and P. Avouris. Electrical conductance of parallel atomic wires. *Physical Review B* 2000; **62**(11): 7235.
39. A. Rochefort and P. Avouris. Electron interference effects on the conductance of doped carbon nanotubes. *Journal of Physical Chemistry* 2000; **104**(44).
40. A. Rochefort, D. R. Salahub and P. Avouris. Effects of finite length on the electronic structure of carbon nanotubes. *Journal of Physical Chemistry B* 1999; **103**, 641–646.
41. R. Martel, H. R. Shea and P. Avouris. Rings of single walled carbon nanotubes. *Nature*, 1999; **398**: 299.
42. R. Martel, T. Schmidt, H. R. Shea, T. Hertel and P. Avouris. Single- and multi-wall carbon nanotube field effect transistors. *Applied Physics Letters* 1998; **73**, 2447.
43. R. Landauer. Spatial variation of currents and fields in localized scatters in metallic conduction. *IBM Journal of Research and Development* 1957; **1**: 223–231.
44. M. Buttiker. *IBM Journal of Research and Development* 1988; **32**: 317.

Index

ESD: Physics and Devices Steven H. Voldman

ISBN: 0-470-84753-0